

84786

From: Chan, Christina
Sent: Tuesday, January 21, 2003 12:07 PM
To: Davis, Minh-Tam; STIC-Biotech/ChemLib
Subject: RE: Rush search request for 09/767215

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
308-3973
CM-1, 9B19

-----Original Message-----

From: Davis, Minh-Tam
Sent: Tuesday, January 21, 2003 11:58 AM
To: Chan, Christina
Subject: Rush search request for 09/767215

Please search in commercial database and in issued patent files:

- 1) Oligomer search for SEQ ID NO:2.
- 2) Amino acids 10-116, 126-420, 568-660, 676-745, 826-1004 of SEQ ID NO:2.

Thank you.

MINH TAM DAVIS
ART UNIT 1642, ROOM 8A01, MB 8E12
305-2008

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

pretty late
2/2000

CRFE

fd no flip
that is even
25 aa as
prim art

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 1/22/03
Date Completed: 1/22/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: 6
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: ②
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:53:00 ; Search time 43 Seconds
(without alignments)
3111.247 Million cell updates/sec

Title: US-09-767-215-2
Perfect score: 1004
Sequence: 1 MGELCRDSALTALDEETLW.....VQIAIDEQKKVWTEQSPR 1004

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1004	100.0	1004	22	AAE07164 Human caspase recr
2	343	34.2	1139	22	AAE07165 Human predicted ca
3	10	1.0	148	22	AAAM24402 Human EST encoded
4	10	1.0	229	20	AAAY73852 Human prostate tum
5	10	1.0	390	20	AAAB82496 Formaldehyde senso
6	10	1.0	1139	22	AAU30127 Novel human secret
7	9	0.9	334	15	AAAR31227 Membrane antigen p
8	9	0.9	348	22	ABBB37003 Peptide #4509 enco
9	9	0.9	348	22	ABBB22317 Protein #4316 enco
10	9	0.9	348	22	AAAM57730 Human brain expres

11	9	0.9	348	22	AAAM70143 Human bone marrow
12	9	0.9	348	22	AAAM17975 Peptide #4409 enco
13	9	0.9	348	22	AAAM30481 Peptide #4518 enco
14	9	0.9	348	22	AAAM05610 Peptide #4232 enco
15	9	0.9	348	23	ABG39782 Human peptide enco
16	9	0.9	599	22	ABBI2297 Human secreted pro
17	9	0.9	599	22	AAAM25255 Human protein sequ
18	9	0.9	1147	22	AAU01207 Human caspase recr
19	9	0.9	1157	22	ABB61471 Drosophila melanog
20	8	0.8	39	21	AAAY66087 BRCAL-associated R
21	8	0.8	48	21	AAAY66089 BRCAL-associated R
22	8	0.8	49	21	AAAY66090 BRCAL-associated R
23	8	0.8	61	21	AAAB42902 Human OREX ORF2666
24	8	0.8	88	23	ABP08766 Human OREX protein
25	8	0.8	106	18	AAAY11170 S. pneumoniae prot
26	8	0.8	106	19	AAAM38530 S. pneumoniae L5SA
27	8	0.8	136	21	AAAG24369 Arabidopsis thalia
28	8	0.8	140	21	AAAG24368 Arabidopsis thalia
29	8	0.8	144	22	AAU59166 Propionibacterium
30	8	0.8	149	21	AAAG48260 Arabidopsis thalia
31	8	0.8	166	21	AAAG48259 Arabidopsis thalia
32	8	0.8	170	23	AAAM48348 Staphylococcus aur
33	8	0.8	197	21	AAAG16552 Arabidopsis thalia
34	8	0.8	250	21	AAAG16551 Arabidopsis thalia
35	8	0.8	274	21	AAAG16550 Arabidopsis thalia
36	8	0.8	274	22	ABG09516 Novel human diagno
37	8	0.8	291	21	AAAG48258 Arabidopsis thalia
38	8	0.8	315	21	AAAG13049 Arabidopsis thalia
39	8	0.8	321	21	AAAG04632 Arabidopsis thalia
40	8	0.8	321	21	AAAG48343 Arabidopsis thalia
41	8	0.8	322	21	AAAB42421 Human OREX ORF2185
42	8	0.8	338	21	AAAG13048 Arabidopsis thalia
43	8	0.8	342	21	AAAG04631 Arabidopsis thalia
44	8	0.8	342	21	AAAG48342 Arabidopsis thalia
45	8	0.8	416	21	AAAG13047 Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAE07164
ID AAE07164 standard: Protein; 1004 AA.

XX AAE07164;

AC AAE07164;

XX 06-NOV-2001 (first entry)

XX Human caspase recruitment domain-14 (CARD-14).

XX Human; caspase recruitment domain-14; CARD-14; chromosome 17;
nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy;
autoimmune disorder; systemic lupus erythematosus; neurological disorder;
Alzheimer's disease; Parkinson's disease; inflammatory disorder; anaemia;
haematological disorder; myelodysplastic syndrome; myocardial infarction;
stroke; immune disorder; Crohn's disease; allergic rhinitis; infection;
cell signalling disorder; cytostatic; immunosuppressive; neurotropic;
neuroprotective; antiviral; antibacterial.

OS Homo sapiens.

Key	Location/Qualifiers
Modified-site	6..9
Domain	/note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
Modified-site	10..116
Modified-site	/label= CARD_domain
Modified-site	12..15
Modified-site	/note= "Casein kinase II phosphorylation site"
Modified-site	18..21
Modified-site	/note= "Casein kinase II phosphorylation site"
Modified-site	25..27
Modified-site	/note= "Protein kinase C phosphorylation site"

FT	Modified-site	60..62	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	91..93	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	114..117	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	117..122	/note= "N-glycosylation site"
FT	Modified-site	121..123	/note= "N-myristoylation site"
FT	Domain	126..420	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	130..135	/label= Coiled_Coil_domain
FT	Modified-site	134..137	/note= "N-myristoylation site"
FT	Modified-site	161..166	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	165..168	/note= "N-myristoylation site"
FT	Modified-site	220..227	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	221..224	/note= "Tyrosine kinase phosphorylation site"
FT	Domain	239..325	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	240..243	/label= k-Box_domain
FT	Modified-site	250..252	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	253..256	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	259..262	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	280..283	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	290..293	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	297..300	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	307..309	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	307..310	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	359..365	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	366..368	/note= "Tyrosine kinase phosphorylation site"
FT	Modified-site	366..369	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	378..381	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	384..386	/note= "Casein kinase II phosphorylation site"
FT	Region	385..406	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	449..452	/note= "Leucine zipper pattern"
FT	Modified-site	453..466	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	463..465	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	470..472	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	501..504	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	511..516	/note= "Casein kinase II phosphorylation site"
FT	Domain	568..660	/note= "N-myristoylation site"
FT	Modified-site	587..592	/label= PD2_domain
FT	Modified-site	589..592	/note= "N-myristoylation site"
FT	Modified-site	602..605	/note= "N-glycosylation site"
FT	Modified-site	634..637	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	653..655	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	674..677	/note= "Protein kinase C phosphorylation site"
FT	Domain	676..745	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	714..719	/label= SH3_domain
FT	Modified-site	725..727	/note= "N-myristoylation site"
FT	Modified-site	725..728	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	733..738	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	737..740	/note= "N-myristoylation site"
FT	Modified-site	759..761	/note= "N-glycosylation site"
FT	Modified-site	760..763	/note= "Protein kinase C phosphorylation site"
FT	Peptide	785..793	/note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT	Modified-site	796..799	/note= "Peroxisomal targeting signal"
FT	Modified-site	800..805	/note= "Casein kinase II phosphorylation site"
FT	Domain	826..1004	/note= "N-myristoylation site"
FT	Modified-site	842..844	/label= Guanylate_kinase_domain
FT	Modified-site	860..863	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	868..870	/note= "Casein kinase II phosphorylation site"
FT	Region	870..872	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	893..896	/note= "RGD cell attachment sequence"
FT	Modified-site	926..929	/note= "Casein kinase II phosphorylation site"
FT	Peptide	941..949	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	944..947	/note= "Peroxisomal targeting signal"
FT	Modified-site	976..979	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	980..985	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	1002..1004	/note= "N-myristoylation site"
FT	Modified-site	1002..1004	/note= "Protein kinase C phosphorylation site"
XX		WO200159065-A2.	
XX		16-AUG-2001.	
XX		22-JAN-2001; 2001WO-US02087.	
XX		09-FEB-2000; 2000US-0181159.	
XX		(MILL-) MILLENNIUM PHARM INC.	
XX		Bertin J;	
XX		WPI; 2001-497073/54.	
XX		N-PSDB; AAD13447.	
XX		An isolated caspase recruitment domain polypeptide useful for	

regulating growth and cell death and useful for the treatment of cancer

Claim 1; Fig 1A-1E; 109pp; English.

The present sequence is human caspase recruitment domain-14 (CARD-14). The CARD-14 gene is located on chromosome 17. The CARD-14 is used for the detection of modulators that modulates the ability of CARD-14 to bind to Bcl-10 and stimulates phosphorylation of Bcl-10 or activation of nuclear factor-kappa B (NF-kB). The CARD-14 is useful for regulating growth and cell death and useful for the treatment of cancer. It is also useful for the treatment of autoimmune disorders (e.g., systemic lupus erythematosus), neurological disorders e.g., Alzheimer's and Parkinson's disease, inflammatory disorders, haematological disorders (e.g., anaemia, myelodysplastic syndromes), myocardial infarctions, strokes, immune disorders (e.g., Crohn's disease, allergic rhinitis), cell signalling disorders and certain viral and bacterial infections.

Sequence 1004 AA;

Query Match	100.0%;	Score 1004;	DB 22;	Length 1004;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1004;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Oy	1	MGELCRDSALTALDEETLWEMESHRRIVRCICPSRLTPYLROAKVLCQDDEEVLHS	60	
Db	1	MGELCRDSALTALDEETLWEMESHRRIVRCICPSRLTPYLROAKVLCQDDEEVLHS	60	
Oy	61	PLRTNSMRAGHLDDLKTRGNKGAIAFLESKLFNPDVYTLVTGLQPDVDFSNFSGLME	120	
Db	61	PLRTNSMRAGHLDDLKTRGNKGAIAFLESKLFNPDVYTLVTGLQPDVDFSNFSGLME	120	
Oy	121	TSKLTECLAGATSGEQEELNKGQVLLRRCQQLQEHGLAETRAEGLHGLEADHSRM	180	
Db	121	TSKLTECLAGATSGEQEELNKGQVLLRRCQQLQEHGLAETRAEGLHGLEADHSRM	180	
Oy	181	KREVAHFVRLKDEMLSLSHYSNALQEKELAAKRSQLEBLYLLKQELQANMVS	240	
Db	181	KREVAHFVRLKDEMLSLSHYSNALQEKELAAKRSQLEBLYLLKQELQANMVS	240	
Oy	241	SCELEQSQSLRTASDQSGDELRNKEKLSLFTSLAEKDILOESLDEARGSRQE	300	
Db	241	SCELEQSQSLRTASDQSGDELRNKEKLSLFTSLAEKDILOESLDEARGSRQE	300	
Oy	301	LVERIHSRLRAVAERQEQWEEKQTLQFQSKMACQLYREKVNALQAVCELOKE	360	
Db	301	LVERIHSRLRAVAERQEQWEEKQTLQFQSKMACQLYREKVNALQAVCELOKE	360	
Oy	361	RQAYSARDSAQRETSQSLVEKDSLRQVFEITDQVCELTQRLQAEPPGVKQEAR	420	
Db	361	RQAYSARDSAQRETSQSLVEKDSLRQVFEITDQVCELTQRLQAEPPGVKQEAR	420	
Oy	421	REPCPREKORLVRMHAIICPRDDSDCLSVSTESQLLSLDSATSSRELVDSPRSSPAPPS	480	
Db	421	REPCPREKORLVRMHAIICPRDDSDCLSVSTESQLLSLDSATSSRELVDSPRSSPAPPS	480	
Oy	481	QOSLYKRYAEDGEPWPFSFSCLEIPEGDPGALPCAKAGDPHLDYELLDADLPQLESSL	540	
Db	481	QOSLYKRYAEDGEPWPFSFSCLEIPEGDPGALPCAKAGDPHLDYELLDADLPQLESSL	540	
Oy	541	QVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEGISVIGNLGTGIFTHRVTP	600	
Db	541	QVSPGRLDVSESGVLMRRRPARRILSQVTMLAFQGDALLEGISVIGNLGTGIFTHRVTP	600	
Oy	601	GSAADQMLRPGCTQIVMDVYEASEPLFKAVLEDDTTLEEAAGLLRRVDGFCCLSVKVNVDG	660	
Db	601	GSAADQMLRPGCTQIVMDVYEASEPLFKAVLEDDTTLEEAAGLLRRVDGFCCLSVKVNVDG	660	
Oy	661	YKRLQLDLEAKVATSGDSFYIRVNLAAMEGRAGKEQLQVHCNEVLHVTDTMFGCGCQWHAHR	720	
Db	661	YKRLQLDLEAKVATSGDSFYIRVNLAAMEGRAGKEQLQVHCNEVLHVTDTMFGCGCQWHAHR	720	
Oy	721	VNSYTMKDTAAHGTIPNYSRAOQQIALIQDMTQOCTVTRKPSGGPKLVRIVSMDKAK	780	

Db	721	VNSYTMKDTAAHGTIPNYSRAOQQIALIQDMTQOCTVTRKPSGGPKLVRIVSMDKAK	780	
Oy	781	ASPLRLSFDRGOLDPSRMGSGSTCFWAEISCLTLVPYTLVWPHRPARPRPVLLVPRAVGKI	840	
Db	781	ASPLRLSFDRGOLDPSRMGSGSTCFWAEISCLTLVPYTLVWPHRPARPRPVLLVPRAVGKI	840	
Oy	841	LSEKLCLLQGFKKCLAEYLSQEEYEAWSQGRDIIQEGEVSGGRCWVTRHAEVSLMEKNTH	900	
Db	841	LSEKLCLLQGFKKCLAEYLSQEEYEAWSQGRDIIQEGEVSGGRCWVTRHAEVSLMEKNTH	900	
Oy	901	ALLDVQLDSVCTLRHMDIPPIVHVSNEKMAKKLKLQRLGTSEEQLLEAARQEGDL	960	
Db	901	ALLDVQLDSVCTLRHMDIPPIVHVSNEKMAKKLKLQRLGTSEEQLLEAARQEGDL	960	
Oy	961	DRAPCLYSSLPADGWSLDGLLSCVRQATADFGKVKVWTEQSPR	1004	
Db	961	DRAPCLYSSLPADGWSLDGLLSCVRQATADFGKVKVWTEQSPR	1004	

RESULT 2

AAE07165

ID AAE07165 standard; Protein; 1139 AA.

XX AAE07165;

XX 06-NOV-2001 (first entry)

XX Human predicted caspase recruitment domain-14 (CARD-14).

XX Human; caspase recruitment domain-14; CARD-14; chromosome 17; nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy; autoimmune disorder; systemic lupus erythematosus; neurological disorder; Alzheimer's disease; Parkinson's disease; inflammatory disorder; anaemia; haematological disorder; myelodysplastic syndrome; myocardial infarction; stroke; immune disorder; Crohn's disease; allergic rhinitis; infection; cell signalling disorder; cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral; antibacterial.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 700

FT /note= "Encoded by TGG"

XX WO200159065-A2.

XX 16-AUG-2001.

XX 22-JAN-2001; 2001WO-US02087.

XX 09-FEB-2000; 2000US-0181159.

XX (MILL-) MILLENNIUM PHARM INC.

XX Bertin J;

XX WPI; 2001-497073/54.

XX N-PSDB; AAD13448.

XX An isolated caspase recruitment domain polypeptide useful for regulating growth and cell death and useful for the treatment of cancer

XX Disclosure; Fig 2A-2C; 109pp; English.

XX The present sequence is predicted human caspase recruitment domain-14 (CARD-14). The CARD-14 gene is located on chromosome 17. The CARD-14 is used for the detection of modulators that modulates the ability of CARD-14 to bind to Bcl-10 and stimulates phosphorylation of Bcl-10 or activation of nuclear factor-kappa B (NF-kB). The CARD-14 is useful for regulating growth and cell death and useful for the treatment of cancer. It is also useful for the treatment of autoimmune disorders (e.g.,

CC systemic lupus erythematosus), neurological disorders e.g., Alzheimer's
CC and Parkinson's disease, inflammatory disorders, haematological disorders
CC (e.g., anaemia, myelodysplastic syndromes), myocardial infarctions,
CC strokes, immune disorders (e.g., Crohn's disease, allergic rhinitis),
CC cell signalling disorders and certain viral and bacterial infections.
XX
SQ Sequence 1139 AA;

Query Match 34.2%; Score 343; DB 22; Length 1139;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 443; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 553 SGVLMRRPARRILSQVYMLAFQGLLEQISVIGNLTGFIHRTVPGSAADQMALRPG 612
DB 603 SGVLMRRPARRILSQVYMLAFQGLLEQISVIGNLTGFIHRTVPGSAADQMALRPG 662
QY 613 TQIVMVDYEAEPLEFKAVLEDTTLEAVGLRRVDFGFCCLSVKVNVDGYKRLQDLEAKV 672
DB 663 TQIVMVDYEAEPLEFKAVLEDTTLEAVGLRRVDFGFCCLSVKVNVDGYKRLQDLEAKV 722
QY 673 ATSGDSFYIRVNLAMEGRAKGLQVHCNEVLHVDTMTFGGCGWHAHRVNSYTKMDTAAH 732
DB 723 ATSGDSFYIRVNLAMEGRAKGLQVHCNEVLHVDTMTFGGCGWHAHRVNSYTKMDTAAH 782
QY 733 GTIPNYSRAQQQIALIQDWTQCTVTRKPSGGPQKLVRIVSMDBKAKASPLRLSFRDQG 792
DB 783 GTIPNYSRAQQQIALIQDWTQCTVTRKPSGGPQKLVRIVSMDBKAKASPLRLSFRDQG 842
QY 793 LDFSRMGSSCTCFWAEBSCLTLPYTLVWPHRPAPRPVLLVPRVAGKILSEKLCILQGFK 852
DB 843 LDFSRMGSSCTCFWAEBSCLTLPYTLVWPHRPAPRPVLLVPRVAGKILSEKLCILQGFK 902
QY 853 KCLAEYLSQEEYEAWSQGRDIIQEGEVSGRCWVTHAVESLMKXNTHALLDVLDSVCT 912
DB 903 KCLAEYLSQEEYEAWSQGRDIIQEGEVSGRCWVTHAVESLMKXNTHALLDVLDSVCT 962
QY 913 LHRMDIFPIVHVSVNEKMAKLLKGLQLRGTSEEOLLEAARQEEGLDRAPCLYSSLAP 972
DB 963 LHRMDIFPIVHVSVNEKMAKLLKGLQLRGTSEEOLLEAARQEEGLDRAPCLYSSLAP 1022
QY 973 DGWSDLDGLLSCVRQAIADQKKV 996
DB 1023 DGWSDLDGLLSCVRQAIADQKKV 1046

RESULT 3
AAM24402
ID AAM24402 standard; Protein; 148 AA.
AC AAM24402;
XX
DT 12-OCT-2001 (first entry)
DE Human EST encoded protein SEQ ID NO: 1927.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX

PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX WPI; 2001-476164/51.
DR N-PSDB; AAH99061.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 20; Page 1236; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SQ Sequence 148 AA;

Query Match 1.0%; Score 10; DB 22; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 823 RPAPRPVLL 832
DB 74 RPAPRPVLL 83

RESULT 4
AAAY73852
ID AAAY73852 standard; Protein; 229 AA.
AC AAAY73852;
XX
DT 14-MAR-2000 (first entry)
DE Human prostate tumor EST fragment derived protein #39.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW treatment.
XX
OS Homo sapiens.
XX
PN DE19820190-A1.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;
XX WPI; 1999-621386/54.
DR N-PSDB; AAZ52870.
XX
PT New human nucleic acid sequences from pancreatic tumors, and related
PT proteins -
XX
PS Claim 23; Page 325; 502pp; German.
XX
CC This invention describes novel polypeptides and their encoding nucleic
CC acids derived from human pancreatic tumor tissue which have cytostatic
CC activity. The sequences are also useful in producing pharmaceutical
CC compositions for treatment of pancreatic tumors. AAAY73814-Y74252
CC represent protein fragments encoded by the human pancreatic tumor cDNA

CC library derived expressed sequence tag (EST) sequences represented in
 CC AA252858-253014.
 XX
 SQ Sequence 229 AA;
 Query Match 1.0%; Score 10; DB 20; Length 229;
 Best Local Similarity 100.0%; Pred. No. 0.65;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 823 RPAPRPVLL 832
 |||||
 Db 77 RPAPRPVLL 86
 |||||
 RESULT 5
 AAB82496
 ID AAB82496 standard; Protein; 390 AA.
 XX
 AC AAB82496;
 XX
 DT 05-SEP-2001 (first entry)
 XX
 DE Formaldehyde sensor protein GfdT.
 XX
 KW Formaldehyde dehydrogenase; repressor; sensor; GfdT;
 XX detection; biosensor; bioremediation; pollutant.
 KW Rhodobacter sphaeroides.
 OS
 XX US6242244-B1.
 PN
 XX
 XX
 PD 05-JUN-2001.
 XX
 XX 16-NOV-1998; 98US-0192983.
 PF
 XX
 PR 28-FEB-1996; 96US-0608241.
 PR 29-AUG-1997; 97US-0919953.
 XX
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Donohue T, Barber R, Withuhn V;
 XX
 XX WPI; 2001-380473/40.
 DR N-PSDB; AAF90616.
 DR
 XX
 PT Oxidation of formaldehyde in a sample by adding a particular bacterial
 PT cell and detecting the decrease in formaldehyde
 XX
 PS Disclosure; Column 31-34; 25pp; English.
 XX
 CC The present sequence is that of the formaldehyde sensor protein
 CC GfdT encoded by an open reading from of the formaldehyde-inducible
 CC regulatory system (see AAF90616) of Rhodobacter sphaeroides.
 CC
 CC Hydrophathy analysis of the sequence suggests GfdT to be a
 CC transmembrane protein. A second open reading frame encodes Gfds
 CC (see AAB82497), a putative soluble protein. By analogy to other
 CC systems, it is thought that a His residue near the C-terminal end
 CC of Gfds may be dephosphorylated when formaldehyde binds to GfdT.
 CC
 CC In the absence of formaldehyde, the phosphate on Gfds may be
 CC transferred to an Asp residue of the repressor GfdR (see AAB82495).
 CC These proteins are utilised in methods of the invention for
 CC detecting and remediating formaldehyde. A claimed method for
 CC oxidizing formaldehyde in a sample involves: (1) providing R.
 CC sphaeroides cells that comprise a glutathione-dependent
 CC formaldehyde dehydrogenase (GSH-FDH) gene, especially the R.
 CC sphaeroides adh1 gene (see AAF90615), an adh1 promoter operably
 CC linked to the GSH-FDH gene, a cis-acting operator, at least 1
 CC protein that senses formaldehyde (especially Gfds and GfdT), and a
 CC trans-acting regulatory protein (especially GfdR) that modulates
 CC transcription of the GSH-FDH gene in response to a signal from the
 CC formaldehyde sensor protein(s), such that the promoter and operator
 CC together specifically direct transcription of the GSH-FDH gene when
 CC the cell is exposed to formaldehyde; (2) adding the bacterial cell

CC to the sample; and (3) detecting a decrease in formaldehyde in the
 CC sample. A biosensor and bioremediation system for environmental
 CC formaldehyde are obtained.
 XX
 SQ Sequence 390 AA;
 Query Match 1.0%; Score 10; DB 22; Length 390;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 308 LRERAVAAER 317
 |||||
 Db 4 LRERAVAAER 13
 |||||
 RESULT 6
 AAU30127
 ID AAU30127 standard; Protein; 1139 AA.
 XX
 AC AAU30127;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #618.
 XX
 DE Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 OS
 PN WO200179449-A2.
 PN
 XX 25-OCT-2001.
 PD
 XX
 PF 16-APR-2001; 2001WO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 XX WPI; 2001-611725/70.
 DR
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 XX
 PS Claim 20; Page 247; 765pp; English.
 XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 XX
 SQ Sequence 1139 AA;
 Query Match 1.0%; Score 10; DB 22; Length 1139;
 Best Local Similarity 100.0%; Pred. No. 2.8;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 PDVDFNFSG 117
| | | | | | | |
Db 108 PDVDFNFSG 117

RESULT 7
AAR51227
ID AAR51227 standard; Protein: 334 AA.

XX AC AAR51227;
XX DT 23-NOV-1994 (first entry)

XX DE Membrane antigen peptide of human sperm.

XX KW Membrane antigen peptide; human sperm; contraceptive vaccine;

XX KW diagnosis of infertility; immunisation.
XX OS Homo sapiens.

XX FH Key Location/Qualifiers
XX FT Peptide 1..335
XX FT /label= membrane antigen peptide of human sperm

XX PN JP06070775-A.
XX PD 15-MAR-1994.

XX PF 27-AUG-1992; 92JP-0228449.
XX PR 27-AUG-1992; 92JP-0228449.

XX PA (ISOJ/) ISOJIMA S.
XX DR WPI; 1994-128676/16.

XX DR N-PSDB; AAQ45175.

XX PT Membrane antigen peptide of human sperm and its coding - used in
XX PT a contraceptive vaccine and for diagnosis of infertility
XX PS Claim 1; Page 8-9; 10pp; Japanese.

XX CC AAR51227 shows the antigen peptide for a monoclonal antibody having
XX CC immobility action of human sperm and inhibitory action of
XX CC fertilisation. It is useful for a contraception vaccine and for the
XX CC diagnosis of fertilisation.
XX SQ Sequence 334 AA;

Query Match 0.9%; Score 9; DB 15; Length 334;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 266 RLKEENEKL 274
| | | | | | | |
Db 55 RLKEENEKL 63

RESULT 8
ABB37003
ID ABB37003 standard; Peptide; 348 AA.

XX AC ABB37003;
XX DT 04-FEB-2002 (first entry)

XX DE Peptide #4509 encoded by human foetal liver single exon probe.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

XX WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human fetal liver -
XX PS Claim 27; SEQ ID NO 29638; 639pp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human foetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC fetal liver. The present sequence is a peptide encoded by a single exon
XX CC nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 348 AA;

Query Match 0.9%; Score 9; DB 22; Length 348;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 266 RLKEENEKL 274
| | | | | | | |
Db 224 RLKEENEKL 232

XX RESULT 9
XX ABB22317
XX ID ABB22317 standard; Protein; 348 AA.

XX AC ABB22317;
XX DT 23-JAN-2002 (first entry)

XX DE Protein #4316 encoded by probe for measuring heart cell gene expression.
XX KW Human; gene expression; heart; microarray; vascular system;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease.
XX OS Homo sapiens.

XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00666.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.

```
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
XX Claim 15; SEQ ID No 24087; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease.
XX hypertension, cardiac arrhythmias and congenital heart disease.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 348 AA;
XX
XX Query Match 0.9%; Score 9; DB 22; Length 348;
XX Best Local Similarity 100.0%; Pred. No. 9.1;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 266 RLKEENEKL 274
XX ||||||||
XX Db 224 RLKEENEKL 232
XX
XX RESULT 10
XX AAM57730
XX ID AAM57730 standard; Protein; 348 AA.
XX
XX AC AAM57730;
XX
XX DT 05-NOV-2001 (first entry)
XX
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 29835.
XX
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer.
XX
XX OS Homo sapiens.
XX
XX PN WO200157275-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00667.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 30449; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
```

```
DR WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 29835; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, of
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention.
XX
XX SQ Sequence 348 AA;
XX
XX Query Match 0.9%; Score 9; DB 22; Length 348;
XX Best Local Similarity 100.0%; Pred. No. 9.1;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 266 RLKEENEKL 274
XX ||||||||
XX Db 224 RLKEENEKL 232
XX
XX RESULT 11
XX AAM70143
XX ID AAM70143 standard; Protein; 348 AA.
XX
XX AC AAM70143;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30449.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00668.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 30449; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
```

```

SQ Sequence 348 AA;
Query Match 0.9%; Score 9; DB 22; Length 348;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 RLKEENEKL 274
| | | | | | | |
Db 224 RLKEENEKL 232

RESULT 12
AAM17975
ID AAM17975 standard; Protein; 348 AA.
XX AC AAM17975;
XX DT 12-OCT-2001 (first entry)
XX DE Peptide #4409 encoded by probe for measuring cervical gene expression.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX OS Homo sapiens.
XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00670.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX Claim 27; SEQ ID No 22801; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
CC (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 348 AA;
Query Match 0.9%; Score 9; DB 22; Length 348;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 RLKEENEKL 274
| | | | | | | |
Db 224 RLKEENEKL 232

RESULT 13
AAM30481
ID AAM30481 standard; Protein; 348 AA.
XX AC AAM30481;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #4518 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX Claim 27; SEQ ID No 30750; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP:
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
SQ Sequence 348 AA;
Query Match 0.9%; Score 9; DB 22; Length 348;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 RLKEENEKL 274
| | | | | | | |
Db 224 RLKEENEKL 232

RESULT 14
AAM05610
ID AAM05610 standard; Protein; 348 AA.
XX AC AAM05610;
XX DT 09-OCT-2001 (first entry)
XX DE Peptide #4292 encoded by probe for measuring breast gene expression.
XX KW Probe; human; breast disease; breast cancer; development disorder;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.

```

XX WO200157270-A2.
PN XX
PD XX
PF 09-AUG-2001.
XX XX
XX 29-JAN-2001; 2001WO-US00661.
XX XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI XX
XX WPI; 2001-476286/51.
DR XX
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
PT XX
XX Claim 27; SEQ ID No 14350; 322pp; English.
PS XX
XX The present invention relates to novel single exon nucleic acid probes
CC (see AAI00010-AA110067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX CC
SQ Sequence 348 AA:
Query Match 0.9%; Score 9; DB 22; Length 348;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 266 RLKEENEKL 274
DB 224 RLKEENEKL 232
|||||||
RESULT 15
ABG39782
ID ABG39782 standard; Peptide; 348 AA.
XX AC ABG39782;
XX XX
XX 19-AUG-2002 (first entry)
XX XX
XX Human peptide encoded by genome-derived single exon probe SEQ ID 29447.
DE XX
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW Chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW primary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW hyaline membrane disease.
XX XX
OS Homo sapiens.
XX XX

WO200186003-A2.
PN XX
PD 15-NOV-2001.
XX XX
XX 30-JAN-2001; 2001WO-US00665.
XX XX
XX 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI XX
XX WPI; 2002-114183/15.
DR XX
XX Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
PT XX
XX Claim 27; SEQ ID No 29447; 634pp; English.
PS XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes. The novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a peptide/protein
CC encoded by a single exon probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX XX
SQ Sequence 348 AA;
Query Match 0.9%; Score 9; DB 23; Length 348;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 RLKEENEKL 274

Db 224 RIKEENEKL 232
|||||

Search completed: January 22, 2003, 08:58:20
Job time : 47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:58:25 ; Search time 19 Seconds
(without alignments)

1554.769 Million cell updates/sec

Title: US-09-767-215-2

Perfect score: 1004

Sequence: 1 MGELCRDRSALTALDETLW.....VROAIADEQKVVWTEQSPR 1004

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	1.0	390	US-09-192-983-5	Sequence 5, Appli
2	8	0.8	106	US-08-858-207A-326	Sequence 326, App
3	8	0.8	446	US-08-874-138-6	Sequence 6, Appli
4	8	0.8	446	US-08-879-941-2	Sequence 2, Appli
5	8	0.8	446	US-09-747-116-2	Sequence 2, Appli
6	7	0.7	24	US-08-851-843A-188	Sequence 188, App
7	7	0.7	24	US-08-974-549A-307	Sequence 307, App
8	7	0.7	24	US-08-854-050-188	Sequence 188, App
9	7	0.7	24	US-09-430-323-188	Sequence 188, App
10	7	0.7	58	PCT-US94-06655-11	Sequence 11, Appli
11	7	0.7	97	US-09-562-737-126	Sequence 126, App
12	7	0.7	105	US-09-395-689-3	Sequence 3, Appli
13	7	0.7	121	US-08-853-659A-36	Sequence 36, Appli
14	7	0.7	136	US-08-946-914-13	Sequence 13, Appli
15	7	0.7	136	US-09-154-750A-79	Sequence 79, Appli
16	7	0.7	136	US-09-656-450-13	Sequence 13, Appli
17	7	0.7	186	US-08-766-551-3	Sequence 3, Appli
18	7	0.7	204	US-09-562-737-127	Sequence 127, App
19	7	0.7	207	US-08-652-877-13	Sequence 13, Appli
20	7	0.7	207	US-08-476-515A-13	Sequence 13, Appli
21	7	0.7	247	US-09-364-230-2	Sequence 2, Appli
22	7	0.7	345	US-09-231-529-1	Sequence 1, Appli
23	7	0.7	345	US-08-977-816-1	Sequence 1, Appli
24	7	0.7	355	US-09-014-969-19	Sequence 19, Appli
25	7	0.7	440	US-09-100-664A-2	Sequence 2, Appli
26	7	0.7	440	US-09-100-664A-3	Sequence 3, Appli
27	7	0.7	440	US-09-100-664A-4	Sequence 4, Appli

28	7	0.7	440	4	US-09-335-983-2	Sequence 2, Appli
29	7	0.7	440	4	US-09-335-983-3	Sequence 3, Appli
30	7	0.7	440	4	US-09-335-983-4	Sequence 4, Appli
31	7	0.7	473	4	US-09-155-036-2	Sequence 2, Appli
32	7	0.7	473	4	US-08-546-793A-2	Sequence 2, Appli
33	7	0.7	474	4	US-09-329-535-2	Sequence 2, Appli
34	7	0.7	474	4	US-09-329-535-9	Sequence 9, Appli
35	7	0.7	495	3	US-08-962-859A-2	Sequence 2, Appli
36	7	0.7	496	4	US-09-155-036-22	Sequence 22, Appli
37	7	0.7	496	4	US-09-155-036-23	Sequence 23, Appli
38	7	0.7	496	4	US-09-155-036-24	Sequence 24, Appli
39	7	0.7	496	4	US-09-155-036-25	Sequence 25, Appli
40	7	0.7	496	4	US-09-155-036-26	Sequence 26, Appli
41	7	0.7	535	4	US-09-137-077-2	Sequence 2, Appli
42	7	0.7	600	4	US-08-836-687B-31	Sequence 31, Appli
43	7	0.7	601	4	US-09-155-036-5	Sequence 5, Appli
44	7	0.7	614	4	US-09-155-036-6	Sequence 6, Appli
45	7	0.7	633	1	US-08-458-477A-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-192-983-5
; Sequence 5, Application US/09192983A
; Patent No. 6242244
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy
; APPLICANT: Barber, Robert
; APPLICANT: Witthuhn, Vernon
; TITLE OF INVENTION: Microbial System for Formaldehyde Sensing and
; FILE REFERENCE: 960296.95505
; CURRENT APPLICATION NUMBER: US/09/192,983A
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/919,953
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 08/608,241
; EARLIER FILING DATE: 1996-02-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-09-192-983-5

Query Match 1.0%; Score 10; DB 4; Length 390;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 L RERAVAAER 317
Db 4 L RERAVAAER 13
|||||||

RESULT 2
US-08-858-207A-326
; Sequence 326, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia

```
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmli, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 326:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
US-08-858-207A-326

Query Match 0.8%; Score 8; DB 4; Length 106;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 135 LOEELNQE 142
Db 52 LOEELNQE 59
|||||||

RESULT 3
US-08-874-138-6
; Sequence 6, Application US/08874138
; Patent No. 5882889
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola
; TITLE OF INVENTION: No. 5882889el Compounds
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/874,138
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: GM10015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-879-941-2

Query Match 0.8%; Score 8; DB 4; Length 446;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 135 LOEELNQE 142
Db 220 LOEELNQE 227
|||||||

RESULT 4
US-08-879-941-2
; Sequence 2, Application US/08879941
; Patent No. 6268172
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola
; TITLE OF INVENTION: NOVEL HISTIDINE KINASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/879,941
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: GM10021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-879-941-2

Query Match 0.8%; Score 8; DB 4; Length 446;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 135 LOEELNQE 142
Db 220 LOEELNQE 227
|||||||

RESULT 5
```

```

US-09-747-116-2
; Sequence 2, Application US/09747116
; Patent No. 6348340
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola
; TITLE OF INVENTION: NOVEL HISTIDINE KINASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09747.116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/879,941
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: GHI0021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-747-116-2
Query Match 0.8%; Score 8; DB 4; Length 446;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 LOEELNQE 142
Db 220 LOEELNQE 227
|||||||

RESULT 6:
US-08-851-843A-188
; Sequence 188, Application US/08851843A
; Patent No. 6093809
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6093809el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:

US-09-747-116-2
Query Match 0.7%; Score 7; DB 3; Length 24;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 TRKPSSG 765
Db 6 TRKPSSG 12
|||||||

RESULT 7:
US-08-974-549A-307
; Sequence 307, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,549A

```

FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 307:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-974-549A-307

Query Match 0.7%; Score 7; DB 4; Length 24;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 TRKPSSG 765
Db 6 TRKPSSG 12

RESULT 8
US-08-854-050-188
Sequence 188, Application US/08854050
Patent No. 6261836
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6261836el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,050
FILING DATE: 09-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 188:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-854-050-188

Query Match 0.7%; Score 7; DB 4; Length 24;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 TRKPSSG 765
Db 6 TRKPSSG 12

RESULT 9
US-09-430-323-188
Sequence 188, Application US/09430323
Patent No. 6309867
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6309867el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco

```
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/430,323
FILING DATE: 29-Oct-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 188:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 188:
US-09-430-323-188
Query Match 0.7%; Score 7; DB 4; Length 24;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 759 TRKPSSG 765
Db 6 TRKPSSG 12
RESULT 10
PCT-US94-06655-11
Sequence 11, Application PC/TUS9406655
GENERAL INFORMATION:
APPLICANT: Chaiken, Irwin
APPLICANT: Graddis, Thomas
APPLICANT: Myszka, David
TITLE OF INVENTION: Coiled-Coil Stem Loop Templates
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: Corporate Patents / P.O. Box 1539
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06655
FILING DATE:
```

```
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sutton, Jeffrey A.
REGISTRATION NUMBER: 34028
REFERENCE/DOCKET NUMBER: P50164
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
TOPOLOGY: circular
MOLECULE TYPE: peptide
PCT-US94-06655-11
Query Match 0.7%; Score 7; DB 5; Length 58;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 900 HALLDVQ 906
Db 10 HALLDVQ 16
RESULT 11
US-09-562-737-126
Sequence 126, Application US/09562737
Patent No. 6428967
GENERAL INFORMATION:
APPLICANT: Herz, Joachim
APPLICANT: Gotthardt, Michael
TITLE OF INVENTION: LDL Receptor Signaling Pathways
FILE REFERENCE: UTSW0708
CURRENT APPLICATION NUMBER: US/09/562,737
NUMBER OF SEQ ID NOS: 132
CURRENT FILING DATE: 2000-05-01
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 126
LENGTH: 97
TYPE: PRT
ORGANISM: human
US-09-562-737-126
Query Match 0.7%; Score 7; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 365 YSARDSA 371
Db 75 YSAROSA 81
RESULT 12
US-09-395-689-3
Sequence 3, Application US/09395689
Patent No. 6387684
GENERAL INFORMATION:
APPLICANT: Hwang, Jaulang
APPLICANT: Hui, cho-Fat
APPLICANT: Chen, Tzong-Yueh
TITLE OF INVENTION: TOPOISOMERASE 1-MEDIATED DNA DELIVERY
FILE REFERENCE: 089191/024001
CURRENT APPLICATION NUMBER: US/09/395,689
CURRENT FILING DATE: 1999-09-13
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 105
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-395-689-3
```

```
Query Match          0.7%; Score 7; DB 4; Length 105;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 KEENEKL 274
DB 58 KEENEKL 64

RESULT 13
US-08-853-659A-36
; Sequence 36, Application US/08853659A
; Patent No. 592522
; GENERAL INFORMATION:
; APPLICANT: Wong, K.K.; Saffer, J.D.
; TITLE OF INVENTION: A Salmonella Sequence, Methods Of Detection
; TITLE OF INVENTION: Of A
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Paul W. Zimmerman
; ADDRESSEE: Intellectual Property Services
; ADDRESSEE: Battelle Memorial Institute
; ADDRESSEE: PNNL P.O. Box 999
; STREET: Washington Way
; CITY: Richland
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 99352
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.50", 1.44 Mb storage
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (Wordperfect 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/853,659A
; FILING DATE: Unknown
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE: N/A
; APPLICATION NUMBER: none
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 121 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-853-659A-36

Query Match          0.7%; Score 7; DB 2; Length 121;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 LSDLSAT 462
DB 65 LSDLSAT 71

RESULT 14
US-08-946-914-13
; Sequence 13, Application US/08946914
; Patent No. 6027916
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
```

```
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,914
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,093
; FILING DATE: 09-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 136 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-946-914-13

Query Match          0.7%; Score 7; DB 3; Length 136;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 904 DVQLDSV 910
DB 127 DVQLDSV 133

RESULT 15
US-09-154-750A-79
; Sequence 79, Application US/09154750A
; Patent No. 6432640
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Polyak, Kornelia
; TITLE OF INVENTION: p53-Induced Apoptosis
; FILE REFERENCE: 1107.75357
; CURRENT APPLICATION NUMBER: US/09/154,750A
; CURRENT FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/059,153
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/079817
; PRIOR FILING DATE: 1998-03-30
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 79
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-154-750A-79

Query Match          0.7%; Score 7; DB 4; Length 136;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 904 DVQLDSV 910
DB 127 DVQLDSV 133

Search completed: January 22, 2003, 09:00:57
Job time : 22 secs
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: January 22, 2003, 08:57:45 ; Search time 14 Seconds
(without alignments)
1447.090 Million cell updates/sec

Title: US-09-767-215-2
Perfect score: 1004
Sequence: 1 MGELCRDSALTALDEETLW.....VRQAIADQKVVWTEQSPR 1004

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 122226 seqs, 20178551 residues

Word size : 0

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1004	100.0	1004	10	US-09-767-215-2
2	343	34.2	1138	10	US-09-767-215-5
3	117	11.7	139	9	US-10-032-159A-16
4	92	9.2	92	9	US-10-032-159A-18
5	9	0.9	348	10	US-09-854-761-37615
6	9	0.9	1247	9	US-10-032-159A-8
7	8	0.8	446	10	US-09-747-116-2
8	7	0.7	24	9	US-09-843-676-188
9	7	0.7	24	9	US-09-438-486-188
10	7	0.7	39	10	US-09-864-761-35935
11	7	0.7	45	10	US-09-864-761-37819
12	7	0.7	61	10	US-09-864-761-33346
13	7	0.7	70	9	US-09-738-626-5287
14	7	0.7	78	10	US-09-815-242-5046
15	7	0.7	92	9	US-10-032-159A-10
16	7	0.7	109	9	US-09-738-626-4070
17	7	0.7	136	10	US-09-728-479-9
18	7	0.7	136	10	US-09-263-689-13
19	7	0.7	186	10	US-09-866-582-26

20	7	0.7	215	9	US-10-001-876-210	Sequence 210, Appl
21	7	0.7	229	10	US-09-767-041-13	Sequence 13, Appl
22	7	0.7	263	10	US-09-925-300-1468	Sequence 1468, Ap
23	7	0.7	303	12	US-10-062-254-326	Sequence 326, App
24	7	0.7	329	10	US-09-880-192-62	Sequence 62, Appl
25	7	0.7	329	10	US-09-788-593A-1	Sequence 1, Appl
26	7	0.7	352	10	US-09-925-299-963	Sequence 963, App
27	7	0.7	361	10	US-09-931-381A-20	Sequence 20, Appl
28	7	0.7	362	9	US-09-898-751A-2	Sequence 2, Appl
29	7	0.7	362	9	US-09-898-751A-4	Sequence 4, Appl
30	7	0.7	362	10	US-09-931-381A-18	Sequence 18, Appl
31	7	0.7	424	9	US-09-738-626-4029	Sequence 4029, Ap
32	7	0.7	440	9	US-10-108-605-199	Sequence 199, App
33	7	0.7	440	9	US-10-108-605-219	Sequence 219, App
34	7	0.7	440	9	US-10-108-605-243	Sequence 243, App
35	7	0.7	461	10	US-09-841-132-399	Sequence 399, App
36	7	0.7	473	9	US-09-796-990-97	Sequence 97, Appl
37	7	0.7	473	9	US-09-796-990-98	Sequence 98, Appl
38	7	0.7	473	9	US-09-796-990-99	Sequence 99, Appl
39	7	0.7	473	9	US-09-796-990-100	Sequence 100, App
40	7	0.7	473	9	US-09-796-990-101	Sequence 101, App
41	7	0.7	473	9	US-09-796-990-102	Sequence 102, App
42	7	0.7	473	9	US-09-796-990-103	Sequence 103, App
43	7	0.7	473	9	US-09-796-990-104	Sequence 104, App
44	7	0.7	473	9	US-09-796-990-105	Sequence 105, App
45	7	0.7	473	9	US-09-796-990-106	Sequence 106, App

ALIGNMENTS

RESULT 1
US-09-767-215-2
; Sequence 2, Application US/09767215
; Patent No. US20020081636A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-142001
; CURRENT APPLICATION NUMBER: US/09/767,215
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/181,159
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1004
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-767-215-2

Query Match	100.0%;	Score	1004;	DB	10;	Length	1004;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	1004;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MGELCRDSALTALDEETLWEMSHRRIVRCIPSRITPYLRQAKVLCOLDEEVLS	60				
Db	1	MGELCRDSALTALDEETLWEMSHRRIVRCIPSRITPYLRQAKVLCOLDEEVLS	60				
QY	61	PLRTNSAMRAGHLDDLKTRGKNGAIAFLSKFHNPDIYTVLTGLQPDVDFSNFSGLM	120				
Db	61	PLRTNSAMRAGHLDDLKTRGKNGAIAFLSKFHNPDIYTVLTGLQPDVDFSNFSGLM	120				
QY	121	TSKLTCLAGATGSGLOEELNQEKQKEVLLRRCCQLQEHGLAETRAEGLHOLEADHSRM	180				
Db	121	TSKLTCLAGATGSGLOEELNQEKQKEVLLRRCCQLQEHGLAETRAEGLHOLEADHSRM	180				
QY	181	KREVSAPHFVRLKDEMLSLSHYSNALQEKELASRCRSLOEELYLKLOELQRANMYS	240				
Db	181	KREVSAPHFVRLKDEMLSLSHYSNALQEKELASRCRSLOEELYLKLOELQRANMYS	240				
QY	241	SCELELQEOSLRTASDQESGDEELNRLKEENEKRLSLTSLAEKDLTLEQSLDEARGSRQ	300				

```
|||||
Db 241 SCELELQEQSLRTASDQESGDEELRLAENEKRLSLFTSLAEKDLILQSLDEARGSRQE 300
QY 301 LVERIHSRLERAVAREQYWEKEQTLLQFQKSKMACQLYREKVNALQAQVCELOKE 360
Db 301 LVERIHSRLERAVAREQYWEKEQTLLQFQKSKMACQLYREKVNALQAQVCELOKE 360
QY 361 RDQAYSARDSAQREYSQSILEKDSLRROVFELTDQVCELRQLTQLOLAQEPGVLKQEAR 420
Db 361 RDQAYSARDSAQREYSQSILEKDSLRROVFELTDQVCELRQLTQLOLAQEPGVLKQEAR 420
QY 421 REPCPREKQRLVRMHAICPRDDSDCSLSVSTESQLLSLDSATSSRELVDSSFRSSPAPPS 480
Db 421 REPCPREKQRLVRMHAICPRDDSDCSLSVSTESQLLSLDSATSSRELVDSSFRSSPAPPS 480
QY 481 QOSLYKRVAEDEGEPWFSFSCLEIPEGDPGALPGAKAGDPHLDYELDLTADLPOLLESSL 540
Db 481 QOSLYKRVAEDEGEPWFSFSCLEIPEGDPGALPGAKAGDPHLDYELDLTADLPOLLESSL 540
QY 541 QPVSGRLDVSSEGLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFTHRVTP 600
Db 541 QPVSGRLDVSSEGLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFTHRVTP 600
QY 601 GSAADQMALRPGCTQIVMVDYEAASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVN 660
Db 601 GSAADQMALRPGCTQIVMVDYEAASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVN 660
QY 661 YKRLLODLEAKVATSGDSFYIRVNLAMEGRAKGELOVHCNEVLHVTDTMFQCGCWHHAHR 720
Db 661 YKRLLODLEAKVATSGDSFYIRVNLAMEGRAKGELOVHCNEVLHVTDTMFQCGCWHHAHR 720
QY 721 VNSYTMKDTAAHGTIPNYSRAOQQLIALIQDMTQOCTVTRKPSGGPQKLVRIVSMKRAK 780
Db 721 VNSYTMKDTAAHGTIPNYSRAOQQLIALIQDMTQOCTVTRKPSGGPQKLVRIVSMKRAK 780
QY 781 ASPLRLSDFRGQLDPSRMEGSSCTCFAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKI 840
Db 781 ASPLRLSDFRGQLDPSRMEGSSCTCFAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKI 840
QY 841 LSEKLCLOGFKKCLAEYLSQBEYEAWSORGDIIOEGEVSGRCWVTRHAVESLMKKNTH 900
Db 841 LSEKLCLOGFKKCLAEYLSQBEYEAWSORGDIIOEGEVSGRCWVTRHAVESLMKKNTH 900
QY 901 ALLDVOLDVSCTLHRMDIFPIVHVSNEKMAKLLKGLQRLGTSEQLLEAARQEEGDL 960
Db 901 ALLDVOLDVSCTLHRMDIFPIVHVSNEKMAKLLKGLQRLGTSEQLLEAARQEEGDL 960
QY 961 DRAPCLYSSLPADGWSDDLGLLSCVROAIADEOKVYVWTEQSPR 1004
Db 961 DRAPCLYSSLPADGWSDDLGLLSCVROAIADEOKVYVWTEQSPR 1004
```

```
RESULT 2
US-09-767-215-5
; Sequence 5, Application US/09767215
; Patent No. US20020081636A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; FILE REFERENCE: 07334-142001
; CURRENT APPLICATION NUMBER: US/09/767,215
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/181,159
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1138
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-767-215-5
```

```
Query Match 34.2%; Score 343; DB 10; Length 1138;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 443; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 553 SGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFTHRVTPGSAADQMALRPG 612
Db 602 SGVLMRRRPARRILSQVTMLAFQGDALLEQISVIGGNLTGIFTHRVTPGSAADQMALRPG 661
QY 613 TQIVMVDYEAASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVN 672
Db 662 TQIVMVDYEAASEPLFKAVLEDTTLEEAVGLLRRVDGFCCLSVKVN 721
QY 673 ATSGDSFYIRVNLAMEGRAKGELOVHCNEVLHVTDTMFQCGCWHHAHRVNSYTMKDTAAH 732
Db 722 ATSGDSFYIRVNLAMEGRAKGELOVHCNEVLHVTDTMFQCGCWHHAHRVNSYTMKDTAAH 781
QY 733 GTIPNYSRAOQQLIALIQDMTQOCTVTRKPSGGPQKLVRIVSMDKAKASPLRLSDFRGQ 792
Db 782 GTIPNYSRAOQQLIALIQDMTQOCTVTRKPSGGPQKLVRIVSMDKAKASPLRLSDFRGQ 841
QY 793 LDPSRMEGSSCTCFAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKLCLOGFK 852
Db 842 LDPSRMEGSSCTCFAESCLTLVPYTLVWPHRPARPRPVLLVPRAVGKILSEKLCLOGFK 901
QY 853 KCLAEYLSQBEYEAWSORGDIIOEGEVSGRCWVTRHAVESLMKKNTHALLDVOLDVSCT 912
Db 902 KCLAEYLSQBEYEAWSORGDIIOEGEVSGRCWVTRHAVESLMKKNTHALLDVOLDVSCT 961
QY 913 LHRMDIFPIVHVSNEKMAKLLKGLQRLGTSEQLLEAARQEEGDLDRAPCLYSSLP 972
Db 962 LHRMDIFPIVHVSNEKMAKLLKGLQRLGTSEQLLEAARQEEGDLDRAPCLYSSLP 1021
QY 973 DGWSDLDGLLSCVROAIADEOKKV 996
Db 1022 DGWSDLDGLLSCVROAIADEOKKV 1045
```

```
RESULT 3
US-10-032-159A-16
; Sequence 16, Application US/10032159A
; Patent No. US20020164703A1
; GENERAL INFORMATION:
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES
; FILE REFERENCE: P-LJ 5100
; CURRENT APPLICATION NUMBER: US/10/032,159A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,457
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 139
; OTHER INFORMATION: Xaa - Any Amino Acid
US-10-032-159A-16
Query Match 11.7%; Score 117; DB 9; Length 139;
Best Local Similarity 100.0%; Pred. No. 1.6e-104;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGELCRDSALTALDEETLWEMSHRRIVRCIPSRILTPYLROAKVLCQDDEEVLHS 60
Db 1 MGELCRDSALTALDEETLWEMSHRRIVRCIPSRILTPYLROAKVLCQDDEEVLHS 60
QY 61 PRLTNSAMRAGHLDDLKTRGKNGAIAFLESUKFHNPDVYTLVTLGLQPDVDFNSFG 117
```

```
Db 61 PRLTNSAMRAGHLLDLLKTRGKNGAIAFLESKFNPDVYTLVTGLQDPVDFSNFSG 117
|||||
RESULT 4
US-10-032-159A-18
; Sequence 18, Application US/10032159A
; Patent No. US20020164703A1
; GENERAL INFORMATION:
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,
; FILE REFERENCE: P-LJ 5100
; CURRENT APPLICATION NUMBER: US/10/032.159A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,457
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-159A-18
Query Match 9.2%; Score 92; DB 9; Length 92;
Best Local Similarity 100.0%; Pred. No. 1e-80;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 EETLWEMESHRRIVRCICPSRLTPYLQAKVLCQDEEVLHSPRLTNSAMRAGHLLD 75
|||||
Db 1 EETLWEMESHRRIVRCICPSRLTPYLQAKVLCQDEEVLHSPRLTNSAMRAGHLLD 60
|||||
Qy 76 LLKTRGKNGAIAFLESKFNPDVYTLVTGLQ 107
|||||
Db 61 LLKTRGKNGAIAFLESKFNPDVYTLVTGLQ 92
|||||

RESULT 5
US-09-864-761-37615
; Sequence 37615, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864.761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37615
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004702.1
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.97
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
; OTHER INFORMATION: EST_HUMAN HIT: AW381570.1, EVALUE 1.00e-89
; OTHER INFORMATION: SWISSPROT HIT: P30622, EVALUE 1.00e-10
; US-09-864-761-37615
Query Match 0.9%; Score 9; DB 10; Length 348;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 266 RLKEENEKL 274
|||||
Db 224 RLKEENEKL 232
|||||

RESULT 6
US-10-032-159A-8
; Sequence 8, Application US/10032159A
; Patent No. US20020164703A1
; GENERAL INFORMATION:
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,
; FILE REFERENCE: P-LJ 5100
; CURRENT APPLICATION NUMBER: US/10/032.159A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,457
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-159A-8
Query Match 0.9%; Score 9; DB 9; Length 1247;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 674 TSGDSFYIR 682
Db 896 TSGDSFYIR 904

RESULT 7
US-09-747-116-2
; Sequence 2, Application US/09747116
; Patent No. US20010006799A1
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola
; TITLE OF INVENTION: NOVEL HISTIDINE KINASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09747.116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/879,941
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: GMI0021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-747-116-2

Query Match 0.8%; Score 8; DB 10; Length 446;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 135 LOEELNQE 142
Db 220 LOEELNQE 227

RESULT 8
US-09-843-676-188
; Sequence 188, Application US/09843676
; Patent No. US20020164786A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: NO. US20020164786A1el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/843.676
; FILING DATE: 26-Apr-2001
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,050
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-0029300S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 188:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 188:
US-09-843-676-188

Query Match 0.7%; Score 7; DB 9; Length 24;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 TRKPSSG 765
Db 6 TRKPSSG 12

RESULT 9
US-09-438-486-188
; Sequence 188, Application US/09438486
; Publication No. US20030009019A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: NO. US20030009019A1el Telomerase
; NUMBER OF SEQUENCES: 223
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/438,486
FILING DATE: 12-NOV-1999
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002931US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 188:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-438-486-188

Query Match 0.7%; Score 7; DB 9; Length 24;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 TRKPSSG 765
|||||
Db 6 TRKPSSG 12

RESULT 10

US-09-864-761-35935
Sequence 35935, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 35935
LENGTH: 39
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010877.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.92
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
OTHER INFORMATION: SWISSPROT HIT: O14607, EVALUE 4.00e-07
OTHER INFORMATION: EST_HUMAN HIT: AA330221.1, EVALUE 7.00e-04
US-09-864-761-35935

Query Match 0.7%; Score 7; DB 10; Length 39;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 267 LKEENEK 273
|||||
Db 17 LKEENEK 23

RESULT 11

US-09-864-761-37819
Sequence 37819, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27

;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 37819
;; LENGTH: 45
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC010736.3
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
;; OTHER INFORMATION: EST_HUMAN HIT: BE542261.1, EVALUATE 3.50e-01
;; OTHER INFORMATION: SWISSPROT HIT: P42917, EVALUATE 2.10e-00
US-09-864-761-37819

Query Match 0.7%; Score 7; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 LQPVSPG 546
Db 20 LQPVSPG 26
|||||||

RESULT 12
US-09-864-761-33346
; Sequence 33346, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmicsa-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04

;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 33346
;; LENGTH: 61
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC002470.5
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.7
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.4
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.9
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.7
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.7
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.6
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.9
;; OTHER INFORMATION: EST_HUMAN HIT: BF381101.1, EVALUATE 3.00e-14
;; OTHER INFORMATION: EST_HUMAN HIT: AW385142.1, EVALUATE 5.20e-02
;; OTHER INFORMATION: SWISSPROT HIT: P41246, EVALUATE 2.00e-00
US-09-864-761-33346

Query Match 0.7%; Score 7; DB 10; Length 61;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 RSSSPAP 478
Db 53 RSSSPAP 59
|||||||

RESULT 13
US-09-738-626-5287
; Sequence 5287, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI

```

; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5287
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5287

Query Match          0.7%: Score 7; DB 9; Length 70;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 210 QEKELAA 216
Db 17 QEKELAA 23

RESULT 14
US-09-815-242-5046
; Sequence 5046, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5046
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae

US-09-815-242-5046
Query Match          0.7%: Score 7; DB 10; Length 78;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 625 PLFKAVL 631
Db 25 PLFKAVL 31

RESULT 15
US-10-032-159A-10
; Sequence 10, Application US/10032159A
; Patent No. US20020164703A1
; GENERAL INFORMATION:
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,
; FILE REFERENCE: P-LJ 5100
; CURRENT APPLICATION NUMBER: US/10/032,159A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,457
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-159A-10

Query Match          0.7%: Score 7; DB 9; Length 92;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 LTPYLRLQ 45
Db 24 LTPYLRLQ 30

Search completed: January 22, 2003, 09:00:30
Job time : 16 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:56:55 ; Search time 26 Seconds
(without alignments)
3712.268 Million cell updates/sec

Title: US-09-767-215-2
Perfect score: 1004
Sequence: 1 MGELRRDSALTALDETLW.....VRQIAIDEQKVKVWTEQSPR 1004

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73:*

1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	1.0	738	2 B69863	two-component sens
2	8	0.8	162	2 B98994	hypothetical prote
3	8	0.8	167	2 H69881	hypothetical prote
4	8	0.8	245	1 XMECNC	DNA replication pr
5	8	0.8	245	2 A31294	chromosome replica
6	8	0.8	270	2 D86135	hypothetical prote
7	8	0.8	324	2 G71061	vegetatable incom
8	8	0.8	332	2 AB3548	hypothetical prote
9	8	0.8	332	2 T26436	hypothetical prote
10	8	0.8	346	2 A12229	cobalamin synthesi
11	8	0.8	348	2 T28806	olfactory receptor
12	8	0.8	366	2 B75481	periplasmic serine
13	8	0.8	381	2 E37665	hypothetical prote
14	8	0.8	388	2 G95177	aminotransferase,
15	8	0.8	388	2 G98043	aminotransferase (
16	8	0.8	418	2 T45807	translation initia
17	8	0.8	446	2 H97929	histidine kinase (
18	8	0.8	662	2 H95934	probable MPAL fami
19	8	0.8	750	2 T21534	hypothetical prote
20	8	0.8	752	1 S40780	translation elonga
21	8	0.8	772	2 T27907	hypothetical prote
22	8	0.8	851	2 T06722	hypothetical prote
23	8	0.8	966	2 S43878	trwC protein - Esc
24	7	0.7	62	2 T30666	hypothetical prote
25	7	0.7	68	1 W2VZA2	7K HindIII-C prote
26	7	0.7	73	2 D86055	hypothetical prote
27	7	0.7	74	2 T36225	hypothetical prote
28	7	0.7	76	2 B91209	hypothetical prote
29	7	0.7	79	2 G95324	probable Trad conf

ALIGNMENTS

RESULT 1

B69863
two-component sensor histidine kinase homolog ykrQ - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: B69863
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchlya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: B69863
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-738 <KUN>
A:Cross-references: GB:Z99111; GB:AL009126; NID:g2633699; PIDN:CAB13226.1; PID:el1849
A:Experimental source: strain 168
C:Genetics:
A:Gene: ykrQ

Query Match 1.0%; Score 10; DB 2; Length 738;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 EELNRLKEEN 271
|||||||
Db 14 EELNRLKEEN 23

RESULT 2

B89894
hypothetical protein SAL058 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: B89894
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: B89894
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-162 <KUR>
A:Cross-references: GB:BA000018; PTD:gl3701015; PIDN:BA42310.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA1058
C:Superfamily: polypeptide deformylase

Query Match 0.8%; Score 8; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 662 KRLQDLE 669
|||||||
DB 29 KRLQDLE 36

RESULT 3
H89881
Hypothetical protein ylxL - Bacillus subtilis
N:Alternate names: sigD 3'-region hypothetical protein C
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: H69881; D55216
R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koether, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
ateuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpatra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: H69881
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-167 <KUN>
A:Cross-references: GB:Z99112; GB:AL009126; NID:g2633902; PIDN:CAB13521.1; PID:el185239;
A:Experimental source: strain 168
R:Marquez-Magana, L.M.; Chamberlin, M.J.
J. Bacteriol. 176, 2427-2434, 1994
A:Title: Characterization of the sigD transcription unit of Bacillus subtilis.
A:Reference number: A55216; MUID:94209246; PMID:8157612
A:Accession: D55216
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-67 <MAR>
A:Cross-references: GB:M20144
C:Genetics:
A:Gene: ylxL

Query Match 0.8%; Score 8; DB 2; Length 167;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 LKEENEKL 274
|||||||
DB 55 LKEENEKL 62

RESULT 4
XMECNC
DNA replication protein dnaC - Escherichia coli (strain K-12)
C:Species: Escherichia coli

C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 01-Mar-2002
C:Accession: A28484; S56588; B25124; C65251
R:Nakayama, N.; Bond, M.W.; Miyajima, A.; Kobori, J.; Arai, K.
J. Biol. Chem. 262, 10475-10480, 1987
A:Title: Structure of Escherichia coli dnaC. Identification of a cysteine residue pos
A:Reference number: A92600; MUID:87280100; PMID:3301836
C:Accession: A28484
A:Molecule type: DNA
A:Residues: 1-245 <NAK>
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from
A:Reference number: S56314; MUID:95334362; PMID:7610040
C:Accession: S56588
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-5, 'D', 7-245 <BUR>
A:Cross-references: EMBL:U14003; NID:gl263172; PIDN:AAA97260.1; PID:g537204
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R:Masai, H.; Bond, M.W.; Arai, K.I.
Proc. Natl. Acad. Sci. U.S.A. 83, 1256-1260, 1986
A:Title: Cloning of the Escherichia coli gene for primosomal protein i: the relations
A:Reference number: A94085; MUID:86149284; PMID:3006041
C:Accession: B25124
A:Molecule type: DNA
A:Residues: 1-14 <MAS>
A:Cross-references: GB:J04030; GB:J02785; GB:M13005; NID:gl45788
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
C:Accession: C65251
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-5, 'D', 7-245 <BLAT>
A:Cross-references: GB:A6000507; GB:U00096; NID:g2367380; PIDN:AC77317.1; PID:gl7908
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: dnaC
A:Map position: 99 min
C:Function:
A:Description: this protein is one of the components of the prepriming protein comple
le-stranded DNA for DNA polymerase reaction
A:Pathway: DNA biosynthesis
C:Superfamily: DNA replication protein dnaC
C:Keywords: DNA biosynthesis; P-loop
F:106-113/Region: nucleotide-binding motif A (P-loop) #status atypical
F:163-169/Region: nucleotide-binding motif B

Query Match 0.8%; Score 8; DB 1; Length 245;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 943 GTSEEQLL 950
|||||||
DB 150 GTSEEQLL 157

RESULT 5
A91294
chromosome replication protein DnaC [Imported] - Escherichia coli (strain O157:H7, su
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 17-May-2002
C:Accession: A91294
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
C:Accession: A91294
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-245 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA838744.1; PID:gl3364799; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: EC5321
C:Superfamily: DNA replication protein dnaC

Query Match 0.8%; Score 8; DB 2; Length 245;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 943 GTSEEQLL 950
|||||
DB 150 GTSEEQLL 157

RESULT 6
DB6135
chromosome replication protein DnaC [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
C:Accession: D86135
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: D86135
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-245 <STO>
A:Cross-references: GB:AB005174; NID:gl2519384; PIDN:AAG59544.1; GSPDB:GN00145; UWGP:Z59
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: dnaC
C:Superfamily: DNA replication protein dnaC

Query Match 0.8%; Score 8; DB 2; Length 245;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 943 GTSEEQLL 950
|||||
DB 150 GTSEEQLL 157

RESULT 7
G71061
hypothetical protein PH1189 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C:Accession: G71061
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohtoku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: G71061
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-270 <KAW>
A:Cross-references: NID:g3236132; PIDN:BAA30289.1; PID:d1031232; PID:g32576
A:Experimental source: strain OF3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1189

Query Match 0.8%; Score 8; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 LKEENEKL 274
|||||

Db 115 LKEENEKL 122
RESULT 8
AB3548
vegetable incompatibility protein het-e-1 [imported] - Brucella melitensis (strain
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AB3548
R:DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A:Reference number: AD3252; PMID:1175668
A:Accession: AB3548
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-324 <KUR>
A:Cross-references: GB:AB008918; PIDN:AAL53549.1; PID:gl7984458; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI10307
A:Map position: II

Query Match 0.8%; Score 8; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 QKEVLLRR 152
|||||
DB 283 QKEVLLRR 290

RESULT 9
T26436
hypothetical protein Y113G7B.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26436
R:Lennard, N.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z20215
A:Accession: T26436
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-332 <WIL>
A:Cross-references: EMBL:AL110477; NID:el542121; PIDN:CAB54324.1; CESP:Y113G7B.1
A:Experimental source: clone Y113G7B
C:Genetics:
A:Gene: CESP:Y113G7B.1
A:Introns: 23/2; 49/2; 259/3

Query Match 0.8%; Score 8; DB 2; Length 332;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 928 NEKMAKKL 935
|||||
DB 7 NEKMAKKL 14

RESULT 10
AT2229
cobalamin synthesis protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AT2229
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 203-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AI2229
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-346 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BAH75091.1; PID:g17132487; GSPDB:GN00179
 A:Experimental source: strain FCC 7120
 C:Genetics:
 A:Gene: cobW
 C:Superfamily: cobW protein

Query Match 0.8%; Score 8; DB 2; Length 346;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 789 DRGOLDPS 796
 |||||
 DB 216 DRGQLDPS 223

RESULT 11
 T28806
 olfactory receptor ODR-10 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T28806
 R:Du, Z.
 submitted to the EMBL Data Library, December 1995
 A:Description: The sequence of C. elegans cosmid C53B7.
 A:Reference number: 220526
 A:Accession: T28806
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-348 <DUF>
 A:Cross-references: EMBL:U42830; PIDN:AAAC48279.1; GSPDB:GN00028; CESP:odr-10
 A:Experimental source: strain Bristol N2; clone C53B7
 C:Genetics:
 A:Gene: CESP:odr-10
 A:Map position: X
 A:Introns: 65/3; 120/2; 174/2; 196/3; 221/3; 263/3; 323/1

Query Match 0.8%; Score 8; DB 2; Length 348;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 LLKTRGKN 83
 |||||
 DB 31 LLKTRGKN 38

RESULT 12
 B75481
 periplasmic serine proteinase, Htra/DegQ/DegS family - Deinococcus radiodurans (strain R
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2002
 C:Accession: B75481
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: B75481
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-366 <WH>
 A:Cross-references: GB:AE001930; GB:AE000513; NID:g6458450; PIDN:AAFL10323.1; PID:g645845
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0745
 A:Map position: 1
 C:Superfamily: Escherichia coli trypsin-like proteinase; GLGF domain homology; trypsin p

Query Match 0.8%; Score 8; DB 2; Length 366;

Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 597 RVTPGSAA 604
 |||||
 DB 274 RVTPGSAA 281

RESULT 13
 S73665
 hypothetical protein P02_orf381 - Mycoplasma pneumoniae (strain ATCC 29342)
 C:Species: Mycoplasma pneumoniae
 A:Variety: ATCC 29342
 C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
 C:Accession: S73665
 R:Himmelreich, R.; Hilbert, H.; Piagens, H.; Pirkil, E.; Li, B.C.; Herrmann, R.
 Nucleic Acids Res. 24, 4420-4449, 1996
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumonia
 A:Reference number: S73327; MUID:97105885; PMID:8948633
 A:Accession: S73665
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-381 <HIM>
 A:Cross-references: EMBL:AE000032; GB:U00089; NID:g1674011; PIDN:AAB95987.1; PID:g167
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C:Genetics:
 A:Genetic code: SGC3

Query Match 0.8%; Score 8; DB 2; Length 381;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 LVSPSTESQ 454
 |||||
 DB 88 LVSPSTESQ 95

RESULT 14
 E95177
 aminotransferase, class II [imported] - Streptococcus pneumoniae (strain TIGR4)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
 C:Accession: E95177
 R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl
 son, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: E95177
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-388 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK75614.1; PID:g14973014; GSPDB:GN00164; TIGR:
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SPI524

Query Match 0.8%; Score 8; DB 2; Length 388;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 SLVEKDSL 385
 |||||
 DB 137 SLVEKDSL 144

RESULT 15
 G98043
 aminotransferase (EC 2.6.1.-) [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001

C;Accession: G98043
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183: 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: G98043
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-388 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAL00180.1; PID:gl5459025; GSPDB:GN00174
C;Genetics:
A:Gene: patB
C;Keywords: aminotransferase

Query Match 0.88; Score 8; DB 2; Length 388;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 SLVEKDSL 385

|||||||

Db 137 SLVEKDSL 144

Search completed: January 22, 2003, 09:00:11
Job time : 30 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:53:45 ; Search time 16 Seconds
(without alignments)

2602.640 Million cell updates/sec

Title: US-09-767-215-2

Perfect score: 1004

Sequence: 1 MGELCRDRSALTALDEETLW.....VRQIADEQKVVWTEQSPR 1004

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	1004	100.0	1004	1	CARE_HUMAN	Q9bx16 homo sapien
2	28	2.8	999	1	CARE_MOUSE	Q99kf0 mus musculu
3	9	0.9	1147	1	CARB_HUMAN	Q9bx17 homo sapien
4	8	0.8	167	1	YLXL_BACSU	P40405 bacillus su
5	8	0.8	245	1	DNAC_ECOLI	P07905 escherichia
6	8	0.8	299	1	HSF6_ARATH	Q9scw4 arabidopsis
7	8	0.8	324	1	PEPL_GADMO	P56272 gadus morhu
8	8	0.8	333	1	PLSX_CLOTS	O65984 clostridium
9	8	0.8	381	1	YF03_MYCPN	P75283 mycoplasma
10	8	0.8	752	1	EFGM_RAT	Q07803 rattus norv
11	7	0.7	68	1	YH07_VACCV	P17357 vaccinia vi
12	7	0.7	88	1	GVPB_BACME	O68677 bacillus me
13	7	0.7	100	1	CHLB_PLESC	P37854 pleurozium
14	7	0.7	100	1	CHLB_POLCU	P37852 polytrichum
15	7	0.7	122	1	RM14_REGAM	O21251 reclinomona
16	7	0.7	135	1	LEG7_HUMAN	P47929 homo sapien
17	7	0.7	142	1	YN77_RHIME	Q9x714 rhizobium m
18	7	0.7	144	1	RPC_BPPHI	P06153 bacterioph
19	7	0.7	182	1	DRTS_PLAVN	P46103 plasmodium
20	7	0.7	186	1	RAYL_HUMAN	Q9bw83 homo sapien
21	7	0.7	186	1	RAYL_MOUSE	Q9d0p8 mus musculu
22	7	0.7	188	1	VATE_THETH	P74901 thermus the
23	7	0.7	201	1	TNFB_MACEU	Q9xt48 macropus eu
24	7	0.7	203	1	RPOC_PROHO	P42075 prochloroth
25	7	0.7	209	1	V281_ARATH	Q9s971 arabidopsis
26	7	0.7	209	1	V282_ARATH	O65421 arabidopsis
27	7	0.7	212	1	HIS1_BACHD	Q9k6z1 bacillus ha
28	7	0.7	231	1	RNH_STRCO	Q9x7r6 streptomyce
29	7	0.7	265	1	JOIN_LYCES	O9fuy6 lycopersico
30	7	0.7	275	1	NFTB_SYAY3	P73451 synecocyst
31	7	0.7	278	1	YKJ5_YEAST	P28320 saccharomyc
32	7	0.7	299	1	BAH_STRHY	Q01109 streptomyce
33	7	0.7	299	1	OIF_BOVIN	P19879 bos taurus

RESULT 1

ID	CARE_HUMAN	STANDARD;	PRT;	1004 AA.
AC	Q9BX16; Q9BVB5;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Caspase recruitment domain protein 14 (CARD-containing MAGUK protein 2) (Carma 2).			
GN	CARD14 OR CARMA2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21192234; PubMed=11278692;			
RA	Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,			
RA	Srinivasula S.M., Merriam S., DiStefano P.S., Alnemri E.S.;			
RT	*CARD11 and CARD14 are novel caspase recruitment domain			
RT	(CARD)/membrane-associated guanylate kinase (MAGUK) family members			
RT	that interact with Bcl10 and activate NF-kappaB.;			
RL	J. Biol. Chem. 276:11877-11882(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21255663; PubMed=11356195;			
RA	Gaide O., Martinon F., Mischeu O., Bonnet D., Thome M., Tschopp J.;			
RT	*Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10			
RT	phosphorylation and NF-kappaB activation.;			
RL	FEBS Lett. 496:121-127(2001).			
RN	[3]			
RP	ERRATUM.			
RA	Gaide O., Martinon F., Mischeu O., Bonnet D., Thome M., Tschopp J.;			
RL	FEBS Lett. 505:198-198(2001).			
RN	[4]			
RP	SEQUENCE OF 1-740 FROM N.A.			
RC	TISSUE=Cervix, and Colon;			
RA	Strausberg R.;			
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.			
CC	!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the			
CC	phosphorylation of Bcl10.			
CC	!- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD			
CC	interaction.			
CC	!- SUBCELLULAR LOCATION: Cytoplasmic.			
CC	!- TISSUE SPECIFICITY: Expressed in placenta. Also detected in HeLa			
CC	S3 cells, but not in the other cancer cell lines tested.			
CC	!- SIMILARITY: CONTAINS 1 CARD DOMAIN.			
CC	!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.			
CC	!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.			
CC	!- CAUTION: Supposed to contain a SH3 domain which is not detected by			
CC	PROSITE, Pfam or SMART.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			

ALIGNMENTS


```

DR EMBL; AF363457; AAK60137.1; -.
DR EMBL; BC004692; AAH04692.1; -.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR PROSITE; PS0209; CARD; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS0052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Coiled coil.
FT DOMAIN 15 107 CARD.
FT DOMAIN 125 411 COILED COIL (POTENTIAL).
FT DOMAIN 572 655 PDZ.
FT DOMAIN 736 986 GUANYLATE KINASE.
FT CONFLICT 736 743 QAQOQLLA -> HLEDHRS (IN REF. 2).
SQ SEQUENCE 999 AA; 113496 MW; D18350DA12430255 CRC64;

Query Match 2.8%; Score 28; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 5.9e-20;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 RGKNGAIAFLESLEKPHNPVYTLVTGLQ 107
      |||||
DB 80 RGKNGAIAFLESLEKPHNPVYTLVTGLQ 107

RESULT 3
ID CARB_HUMAN STANDARD; PRT; 1147 AA.
AC Q9BXL7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 11 (CARD-containing MAGUK protein
DE 3) (Carna 1).
GN CARD11 OR CARMA1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21192234; PubMed=11278692;
RA Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA Srinivasula S.M., Merriam S., Distefano P.S., Alnemri E.S.;
FT "CARD11 and CARD14 are novel caspase recruitment domain
FT (CARD)/membrane-associated guanylate kinase (MAGUK) family members
FT that interact with Bcl10 and activate NF-kappaB."
FT J. Biol. Chem. 276:11877-11882(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21255663; PubMed=11356195;
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschoopp J.;
RT "Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10
RT phosphorylation and NF-kappaB activation.";
RL FEBS Lett. 496:121-127(2001).
RN [3]
RP ERATUM.
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschoopp J.;
RL FEBS Lett. 505:198-198(2001).
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the
CC phosphorylation of Bcl10.
CC -!- SUBUNIT: CARD11 and Bcl10 bind to each other by CARD-CARD
CC interaction.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Detected in adult peripheral blood leukocytes,
CC thymus, spleen and liver. Also found in promyelocytic leukemia HL-
CC 60 cells, chronic myelogenous leukemia K562 cells, Burkitt's
CC lymphoma Raji cells and colorectal adenocarcinoma SW480 cells. Not
CC detected in HeLa S3, Molt-4, A549 and G431 cells.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.

```

-!- CAUTION: Supposed to contain a SH3 domain which is not detected by PROSITE, Pfam or SMART.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF322641; AAG53402.1; -.
 Genew; HGNC:16393; CARD11.
 InterPro; IPR000619; Guanylate_kin.
 InterPro; IPR001478; PDZ.
 SMART; SM00228; PDZ; 1.
 PROSITE; PS0209; CARD; 1.
 PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
 PROSITE; PS0052; GUANYLATE_KINASE_2; FALSE_NEG.
 PROSITE; PS0106; PDZ; FALSE_NEG.
 KW Coiled coil.

FT DOMAIN 11 103 CARD.
 FT DOMAIN 123 442 COILED COIL (POTENTIAL).
 FT DOMAIN 673 748 PDZ.
 FT DOMAIN 966 1133 GUANYLATE KINASE.
 FT CONFLICT 808 808 P -> L (IN REF. 2).
 SQ SEQUENCE 1147 AA; 132641 MW; 913A4B015D2B36CC CRC64;

Query Match 0.9%; Score 9; DB 1; Length 1147;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 674 TSGDSFYIR 682
 DB 764 TSGDSFYIR 772

RESULT 4
 YLXL_BACSU STANDARD; PRT; 167 AA.
 ID YLXL_BACSU STANDARD; PRT; 167 AA.
 AC P40405;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ylxL precursor.
 GN YLXL.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,

RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yoshikawa K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256(1997).
RL [2]
RP SEQUENCE OF 1-67 FROM N.A.
RX MEDLINE=88169477; PubMed=2832368;
RA Helmann J.D., Marquez L.M., Chamberlin M.J.;
RA "Cloning, sequencing, and disruption of the *Bacillus subtilis* sigma
RT 28 gene.";
RL J. Bacteriol. 170:1568-1574(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z99112; CAB13521.1; -;
DR EMBL; M20144; -; NOT_ANNOTATED_CDS.
DR Subtilisin; BG10752; YlxH.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 167 HYPOTHETICAL PROTEIN YLXL.
SQ SEQUENCE 167 AA; 19119 MW; 356021A73B679B69 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 167;

Best Local Similarity 100.0%; Pred. No. 4.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 LKEENEKL 274

DB 55 LKEENEKL 62

RESULT 5

DNAC_ECOLI

ID DNAC_ECOLI STANDARD; PRT; 245 AA.

AC P07905;

DT 01-AUG-1988 (Rel. 08, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE DNA replication protein dnaC.

GN DNAC OR DNAD OR B4361 OR Z5961 OR ECS5321.

OS *Escherichia coli*, and

OS *Escherichia coli* O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC *Escherichia*.

OX NCBI_TaxID=562, 83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=87280100; PubMed=3301836;

RA Nakayama N., Bond M.W., Miyajima A., Kobori J., Arai K.-I.;

RT "Structure of *Escherichia coli* dnaC. Identification of a cysteine

RT residue possibly involved in association with dnaB protein.";

RL J. Biol. Chem. 262:10475-10480(1987).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=89008392; PubMed=2844800;

RA Masai H., Arai K.-I.;

RT "Operon structure of *dnaT* and *dnaC* genes essential for normal and

RT stable DNA replication of *Escherichia coli* chromosome.";

J. Biol. Chem. 263:15083-15093(1988).

RL [3]

RP SEQUENCE FROM N.A.

RX STRAIN=K12 / MG1655;

RC MEDLINE=95334362; PubMed=7610040;

RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,

RA Blattner F.R.;

RT "Analysis of the *Escherichia coli* genome VI: DNA sequence of the

RT region from 92.8 through 100 minutes.";

RL Nucleic Acids Res. 23:2105-2119(1995).

RL [4]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";

RL Nature 409:529-533(2001).

RL [5]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA Ikeda T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,

RA Kihara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*

RT O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001).

RL [6]

RP SEQUENCE OF 1-14 FROM N.A.

RC STRAIN=K12;

RX MEDLINE=86149284; PubMed=3006041;

RA Masai H., Bond M.W., Arai K.-I.;

RT "Cloning of the *Escherichia coli* gene for primosomal protein i: the

RT relationship to *dnaT*, essential for chromosomal DNA replication.";

RL Proc. Natl. Acad. Sci. U.S.A. 83:1256-1260(1986).

CC -!- FUNCTION: THIS PROTEIN IS REQUIRED FOR CHROMOSOMAL REPLICATION. IT

CC FORMS, IN CONCERT WITH DNAB PROTEIN AND OTHER PREPRIMING PROTEINS

CC DNAT, N, N', N'' A PREPRIMING PROTEIN COMPLEX ON THE SPECIFIC SITE

CC OF THE TEMPLATE DNA RECOGNIZED BY PROTEIN N'.

CC -!- SIMILARITY: BELONGS TO THE DNAC FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; J04030; AAA23700.1; -;

DR EMBL; U14003; AAA97260.1; -;

DR EMBL; AE000507; AAC77317.1; -;

DR EMBL; AE005667; AAG59544.1; -;

DR EMBL; AP002569; BAB38744.1; -;

DR PIR; A28484; XMECNC.

DR PIR; B25124; B25124.

DR ECO2DBASE; I030.3; 6TH EDITION.

DR EcoGene; EG10237; dnaC.

DR InterPro; IPR003593; AAA_ATPase.

DR SMART; SM00382; AAA; 1.

KW DNA replication; Primosome; Complete proteome.

FT SITE 69 69 PROBABLY INVOLVED IN THE INTERACTION

FT CONFLICT 6 6 WITH THE DNAB PROTEIN.

FT SEQUENCE 245 AA; 27935 MW; C5CC7232E21F1F1 CRC64;

Query Match

0.8%; Score 8; DB 1; Length 245;

```
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 943 GTSEBQLL 950
DB 150 GTSEBQLL 157

RESULT 6
HSP6_ARATH
ID HSP6_ARATH STANDARD; PRT; 299 AA.
AC Q9SCW4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Heat shock factor protein 6 (HSF 6) (Heat shock transcription factor
6) (HSF 6).
GN HSF6 OR AT5G62020 OR MTG10.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Green siliques;
RA Schoeffl F., Praendl R.;
RT "De-repression of heat shock protein synthesis in transgenic plants.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:297-308(1998).
CC -!- FUNCTION: DNA-BINDING PROTEIN THAT SPECIFICALLY BINDS HEAT SHOCK
CC PROMOTER ELEMENTS (HSE) AND ACTIVATES TRANSCRIPTION (BY
CC SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- PTM: EXHIBITS TEMPERATURE-DEPENDENT PHOSPHORYLATION
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HSF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ251867; CAB63802.1; -
DR EMBL: AB016880; BAB10163.1; -
DR HSP: P22813; 1HKT.
DR InterPro: IPR000232; HSF_DNA_bind.
DR InterPro: IPR002341; HSF_ETFS.
DR Pfam: PF00447; HSF_DNA_bind; 1.
DR PRINTS: PR00056; HSFDOMAIN.
DR ProDom: PD001788; HSF_DNA_bind; 1.
DR SMART: SM00415; HSF; 1.
DR PROSITE: PS00434; HSF_DOMAIN; 1.
KW Transcription regulation; Nuclear protein; DNA-binding; Activator;
KW Phosphorylation; Heat shock; Multigene family.
FT DNA_BIND 21 115 BY SIMILARITY.
FT DOMAIN 145 148 POLY-ASN.
FT DOMAIN 240 248 POLY-GLU.
SQ SEQUENCE 299 AA; 34051 MW; 04D86398693B1997 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 299;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 EENEKLS 276
DB 177 EENEKLS 184

RESULT 7
PEP1_GADMO
ID PEP1_GADMO STANDARD; PRT; 324 AA.
AC P56272;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pepsin IIB (EC 3.4.23.-).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE, AND X-RAY CRYSTALLOGRAPHY.
RC TISSUE=Stomach;
RA Karlsten S., Hough E., Olsen R.L.;
RT "Structure and proposed amino-acid sequence of a pepsin from Atlantic
RT cod (Gadus morhua)".
RL Acta Crystallogr. D 54:32-46(1998).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
DR PDB: 1AM5; 24-DEC-97.
DR InterPro: IPR001461; AspproteaseA1.
DR InterPro: IPR001969; Aspprotease_site.
DR Pfam: PF00026; asp; 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; 3D-structure.
FT ACT_SITE 32 32 BY SIMILARITY.
FT ACT_SITE 214 214 BY SIMILARITY.
FT DISULFID 45 50 BY SIMILARITY.
FT DISULFID 206 209 BY SIMILARITY.
FT DISULFID 247 280 BY SIMILARITY.
SQ SEQUENCE 324 AA; 34014 MW; EE3A6097B6941DD7 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 324;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 376 SQSLVEKD 383
DB 142 SQSLVEKD 149

RESULT 8
PLSX_CLOTS
ID PLSX_CLOTS STANDARD; PRT; 333 AA.
AC O65984;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fatty acid/phospholipid synthesis protein plsx.
GN PLSX.
OS Clostridium thermosaccharolyticum (Thermoanaerobacterium
OS thermosaccharolyticum).
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacterium.
OX NCBI_TaxID=1517;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 571;
RA van Rinsum A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NOT KNOWN, PROBABLY INVOLVED IN FATTY ACID OR
```

```

CC PHOSPHOLIPID SYNTHESIS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PLSX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ004870; CAU06178.1; ALT_INIT.
DR InterPro; IPR003664; FA_synthesis.
DR Pfam; PF02504; FA_synthesis; 1.
DR ProDom; PD006974; FA_synthesis; 1.
DR TIGRFAMs; TIGR00182; plsx; 1.
KW Fatty acid biosynthesis; Phospholipid biosynthesis.
SQ SEQUENCE 333 AA; 35930 MW; 8A932B1A96D88EC9 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 LKQELQR 235
DB 241 LKQELQR 248

RESULT 9
YF03_MYCPN
ID YF03_MYCPN STANDARD; PRT; 381 AA.
AC P75283;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Very hypothetical mpcp-like protein MPN503 (P02_orf381).
GN MPN503 OR MP339.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SIMILARITY: BELONGS TO THE MGPC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000032; AAB95987.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 381 AA; 40896 MW; 5CFD05505E344852 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 LVSSTESQ 454
DB 88 LVSSTESQ 95

RESULT 10

```

```

EFGM_RAT
ID EFGM_RAT STANDARD; PRT; 752 AA.
AC Q07803;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Elongation factor G, mitochondrial precursor (MEF-G).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93324327; PubMed=8332461;
RA Barker C.S., Makris A., Patriotis C., Bear S.E., Tschlis P.N.;
RT "Identification of the gene encoding the mitochondrial elongation
RT factor G in mammals."
RL Nucleic Acids Res. 21:2641-2647(1993).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC RIBOSOME.
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC -!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES WITH THE HIGHEST LEVEL
CC IN LIVER, THYMUS AND BRAIN.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L14684; AAA1107.1;
DR PIR; S40780; S40780.
DR HSSP; P13551; IFNM.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000640; EFG_C.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF-GTPbind.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR TIGRFAMs; TIGR00484; EF-G; 1.
DR PROSITE; PS00301; EFATOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; Mitochondrion;
KW Transit peptide; GTP-binding.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 752 ELONGATION FACTOR G.
FT NP_BIND 54 61 GTP (BY SIMILARITY).
FT NP_BIND 121 125 GTP (BY SIMILARITY).
FT NP_BIND 175 178 GTP (BY SIMILARITY).
SQ SEQUENCE 752 AA; 83770 MW; DFB6108D38A72E4 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 752;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 429 QRLVRMHA 436
DB 396 QRLVRMHA 403

RESULT 11
YH07_VACCV
ID YH07_VACCV STANDARD; PRT; 68 AA.
AC P17357;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)

```

```
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Hypothetical 7.8 kDa HINDIII-C protein.
OS Vaccinia virus (strain WR).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89073756; PubMed=2849238;
RA Kotwal G.J., Moss B.;
RT "Analysis of a large cluster of nonessential genes deleted from a
RL vaccinia virus terminal transposition mutant.";
RL Virology 167:524-537(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22812; AAA69593.1; -.
DR PIR; B31829; WZVZA2.
KW Hypothetical protein; Early protein.
SQ SEQUENCE 68 AA; 7814 MW; 47DF7558D134D949 CRC64;

Query Match          0.7%; Score 7; DB 1; Length 68;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 VLHVTD 708
DB 36 VLHVTD 42
|||||||

RESULT 12
ID GVPE_BACME STANDARD; PRT; 88 AA.
AC O68677;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Gas vesicle structural protein B (GVP B).
GN GVPB.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1404;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VT1660;
RX MEDLINE=98233742; PubMed=9573198;
RA Li N., Cannon M.C.;
RT "Gas vesicle genes identified in Bacillus megaterium and functional
RL expression in Escherichia coli.";
RL J. Bacteriol. 180:2450-2458(1998).
CC -!- FUNCTION: Gas vesicles are small, hollow, gas filled protein
CC structures that are found in several microbial planktonic
CC microorganisms. They allow the positioning of the organism at
CC the favorable depth for growth. GvpA type proteins form the
CC essential core of the structure.
CC -!- SUBCELLULAR LOCATION: GAS VESICLE MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE GAS VESICLE PROTEIN TYPE A FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF053765; AAC38416.1; -.

```

```
DR InterPro; IPR000638; Gas_vesicle.
DR Pfam; PF00741; Gas_vesicle; 1.
DR ProDom; PD003598; Gas_vesicle; 1.
DR PROSITE; PS00234; GAS_VESICLE_A_1; 1.
DR PROSITE; PS00669; GAS_VESICLE_A_2; 1.
KW Gas vesicle.
SQ SEQUENCE 88 AA; 9618 MW; 5F089DE77358D84A CRC64;

Query Match          0.7%; Score 7; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 638 EAVGLLR 644
DB 59 EAVGLLR 65
|||||||

RESULT 13
ID CHLB_PLESC STANDARD; PRT; 100 AA.
AC P37854;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Light-independent protochlorophyllide reductase subunit B
DE (EC 1.18.-.-) (LI-POR subunit B) (DPOB subunit B) (Fragment).
GN CHLB.
OS Pleurozium schreberi (Moss).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Bryopsida; Bryidae; Hypnales; Hylocomiaceae; Pleurozium.
OX NCBI_TaxID=34163;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96428856; PubMed=8812302;
RA Boivin R., Richard M., Beauseigle D., Bousquet J., Bellemare G.;
RT "Phylogenetic inferences from chloroplast chlB gene sequences of
RT Nephrolepis exaltata (Filicopsida), Ephedra altissima (Gnetopsida),
RT and diverse land plants.";
RL Mol. Phylogenet. Evol. 6:19-29(1996).
CC -!- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of
CC protochlorophyllide (Pchlide) to form chlorophyllide a (Chlide)
CC (By similarity). This reaction is light-independent.
CC -!- PATHWAY: Light-independent chlorophyll biosynthesis.
CC -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
CC of three subunits; chlL, chlN and chlB. Could form a
CC heterotetramer of two chlB and two chlN subunits (By similarity).
CC -!- SIMILARITY: BELONGS TO THE CHLB / BCHB / BCHZ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L25774; AAC37495.1; -.
DR InterPro; IPR000510; Oxred_nitrognasel.
DR Pfam; PF00148; oxidored_nitro; 1.
KW Oxidoreductase; Photosynthesis; Chlorophyll biosynthesis; Chloroplast.
FT NON_TER 1
FT NON_TER 100
SQ SEQUENCE 100 AA; 11659 MW; 40B101BE6D66B5E0 CRC64;

Query Match          0.7%; Score 7; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 662 KRLLQDL 668
DB 1 KRLLQDL 7
|||||||

```

```

RESULT 14
CHLB_POLCU
ID CHLB_POLCU STANDARD; PRT; 100 AA.
AC P37852;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Light-independent protochlorophyllide reductase subunit B
DE (EC 1.18.-.-) (LI-POR subunit B) (DPOR subunit B) (Fragment).
CN CHLB.
OS Polytrichum commune (Moss).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Bryophyta;
OC Polytrichopsida; Polytrichales; Polytrichaceae; Polytrichum.
OX NCBI_TaxID=3213;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96426856; PubMed=8812302;
RA Boivin R., Richard M., Beauseigle D., Bousquet J., Bellemare G.;
RT "Phylogenetic inferences from chloroplast chlB gene sequences of
RT Nephrolepis exaltata (Filicopsida), Ephedra altissima (Gnetopsida),
RT and diverse land plants.";
RL Mol. Phylogenet. Evol. 6:19-29(1996).
CC -!- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of
CC protochlorophyllide (PchlId) to form chlorophyllide a (ChlId)
CC (By similarity). This reaction is light-independent.
CC -!- PATHWAY: Light-independent chlorophyll biosynthesis.
CC -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
CC of three subunits: chlL, chlN and chlB. Could form a
CC heterotetramer of two chlB and two chlN subunits (By similarity).
CC -!- SIMILARITY: BELONGS TO THE CHLB / BCHB / BCHZ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L25772; AAC37494.1; -
DR InterPro; IPR000510; Oxred_nitrognsl.
DR Pfam; PF00148; oxidored_nitro; 1.
KW Oxidoreductase; Photosynthesis; Chlorophyll biosynthesis; Chloroplast.
FT NON_TER 1
FT NON_TER 100
SQ SEQUENCE 100 AA; 11573 MW; 984F131F432400A3 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 662 KRLQLD 568
Db 1 KRLQLD 7

RESULT 15
RM14_RECAM
ID RM14_RECAM STANDARD; PRT; 122 AA.
AC O21251;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Mitochondrial 60S ribosomal protein L14.
GN RPL14.
OS Reclinomonas americana.
OC Reclinomonadion.
OC Eukaryota; core jakobids; Reclinomonas.
OX NCBI_TaxID=48483;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97311393; PubMed=9168110;
RA Lang B.F., Burger G., O'Kelly C.J., Cedergren R., Golding G.B.,
RA Lemieux C., Sankoff D., Turmel M., Gray M.W.;
RT "An ancestral mitochondrial DNA resembling a eubacterial genome in
RT miniature.";
RL Nature 387:493-497(1997).
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC -!- SIMILARITY: BELONGS TO THE L14P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF007261; AAD11878.1; -
DR HSSP; P04450; 1WHI.
DR InterPro; IPR000218; Ribosomal_L14.
DR Pfam; PF00238; Ribosomal_L14; 1.
DR ProDom; PD001093; Ribosomal_L14; 1.
DR TIGRfams; TIGR01067; rplN_bact; 1.
DR PROSITE; PS00049; RIBOSOMAL_L14; 1.
KW Ribosomal protein; Mitochondrion.
SQ SEQUENCE 122 AA; 13415 MW; F942C64410D4EE53 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 122;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 769 KLVRIVS 775
Db 110 KLVRIVS 116

Search completed: January 22, 2003, 08:58:42
Job time : 20 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:54:50 ; Search time 48 Seconds
(without alignments)
4309.819 Million cell updates/sec

Title: US-09-767-215-2

Perfect score: 1004

Sequence: 1 MGELCRRDSALTALDETLW.....VQIAIDEQKVKVWTEQSPR 1004

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	1.0	390	2 Q8VVZ7	Q8vvz7 rhodobacter
2	10	1.0	738	16 Q31661	Q31661 bacillus su
3	9	0.9	435	5 Q9NF14	Q9nft4 trichomonas
4	9	0.9	512	10 Q9FRY1	Q9fr1 ceratopteris
5	9	0.9	759	4 Q9HCQ3	Q9hcg3 homo sapien
6	9	0.9	949	5 Q9SU22	Q9su22 drosophila
7	9	0.9	949	5 P91635	P91635 drosophila
8	9	0.9	1020	4 Q8WU84	Q8wu84 homo sapien
9	9	0.9	1157	5 Q9WIM7	Q9wim7 drosophila
10	9	0.9	1171	4 Q8TES3	Q8tes3 homo sapien
11	8	0.8	143	4 Q9H604	Q9h604 homo sapien
12	8	0.8	157	17 Q8THL4	Q8thl4 methanosarc
13	8	0.8	162	16 Q9UQ03	Q9uq3 staphylococ
14	8	0.8	223	2 Q9AN11	Q9an11 bradyrhizob
15	8	0.8	262	10 Q9FSR5	Q9fsr5 oryza sativ
16	8	0.8	270	17 Q58907	Q58907 pyrococcus

17	8	0.8	301	10 Q9FTG5	Q9fig5 arabidopsis
18	8	0.8	319	10 Q9ZSE0	Q9zse0 pinus radia
19	8	0.8	324	16 Q8YD70	Q8y70 brucella me
20	8	0.8	332	5 Q9U2Y1	Q9u2y1 caenorhabdi
21	8	0.8	339	5 Q18807	Q18807 caenorhabdi
22	8	0.8	346	16 Q8YRQ2	Q8yrq2 anabaena sp
23	8	0.8	366	16 Q9RWC4	Q9rwc4 deinococcus
24	8	0.8	388	16 Q97PS5	Q97ps5 streptococc
25	8	0.8	418	4 Q9H6G7	Q9h6g7 homo sapien
26	8	0.8	418	10 Q9M218	Q9m218 arabidopsis
27	8	0.8	420	11 Q60552	Q60552 mesocricetu
28	8	0.8	435	3 Q8X0E6	Q8x0e6 neurospora
29	8	0.8	441	10 Q9M4T7	Q9m4t7 arabidopsis
30	8	0.8	441	10 Q9C523	Q9c523 arabidopsis
31	8	0.8	446	2 Q9F2F5	Q9f2f5 streptococc
32	8	0.8	454	2 Q8RSB5	Q8rsb5 halomonas c
33	8	0.8	501	5 Q9BKQ7	Q9bkq7 caenorhabdi
34	8	0.8	571	10 Q43060	Q43060 porphyridiu
35	8	0.8	571	10 Q43061	Q43061 porphyridiu
36	8	0.8	640	10 Q9FHK7	Q9fhk7 arabidopsis
37	8	0.8	662	16 Q92VG1	Q92vg1 rhizobium m
38	8	0.8	673	4 Q8XV3	Q8xv3 homo sapien
39	8	0.8	693	16 Q8XV01	Q8xvq1 raistonia s
40	8	0.8	750	5 Q9XV52	Q9xv52 caenorhabdi
41	8	0.8	751	11 Q924I0	Q924i0 mus musculu
42	8	0.8	751	11 Q921D6	Q921d6 mus musculu
43	8	0.8	777	5 Q23529	Q23529 caenorhabdi
44	8	0.8	808	4 Q15083	Q15083 homo sapien
45	8	0.8	841	16 Q9ACU8	Q9acu8 streptomyce

ALIGNMENTS

RESULT 1

Q8VVZ7 ID Q8VVZ7 PRELIMINARY; PRT; 390 AA.
 AC Q8VVZ7;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE GfDt.
 GN GfDt.
 OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 CC Rhodobacter.
 OX NCBI_taxid=1063;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Witthuhn V.C. Jr., Donohue T.J.;
 RT "GfDTS negatively regulates transcription of Rhodobacter sphaeroides
 gene products involved in formaldehyde oxidation.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF435819; AAL30776.1; -
 SQ SEQUENCE 390 AA: 41929 MW; BDA7AC33D3F5AE84 CRC64;

Query Match 1.0%; Score 10; DB 2; Length 390;

Best Local Similarity 100.0%; Pred. No. 0.33; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0;

QY 308 LRERAVAAER 317

Db 4 LRERAVAAER 13

RESULT 2

Q31661 ID Q31661 PRELIMINARY; PRT; 738 AA.
 AC Q31661;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE YKQ protein.

GN YKRO.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Bacillaceae; Bacillus.
 ON NCBI_TaxID=1423;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=168:
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Gulseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maueil C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale F.,
 RA Sato T., Scanlan E., Schlecht S., Schroter P., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognotti A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzeneberger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 subtilis";
 RL Nature 390:249-256(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168:
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.
 DR EMBL; Z99111; CAB13226.1; -.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR004358; Bact_sens_pr_C.
 DR InterPro; IPR003661; His_kinA.
 DR InterPro; IPR004359; HIS_KIN_sig.
 DR InterPro; IPR000014; PAS_domain.
 DR Pfam; PF02518; HATPase_c; 1.
 DR Pfam; PF00989; PAS; 3.
 DR Pfam; PF00512; signal; 1.
 DR PRINTS; PR00344; BCTRLSNSOR..
 DR SMART; SM00387; HATPase_c; 1.
 DR SMART; SM00388; HSKA; 1.
 DR SMART; SM00091; PAS; 4.
 DR TIGRFAWS; TIGR00229; sensory_box; 4.
 KW Kinase; Phosphorylation; Sensory transduction; Transferase;
 KW Complete proteome.
 SQ SEQUENCE 738 AA; 85513 MW; FE27DAB7F08AC64B CRC64;

Query Match 1.0%; Score 10; DB 16; Length 738;
 Best Local Similarity 100.0%; Pred. No. 0.58;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 EELNRLKEEN 271
 |||
 DB 14 EELNRLKEEN 23

RESULT 3
 Q9NFT4 PRELIMINARY; PRT; 435 AA.
 AC Q9NFT4; 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Coronin.
 GN COR1.
 OS Trichomonas vaginalis.
 OC Eukaryota; Parabasalidea; Trichomonadida; Trichomonadidae;
 OC Trichomonas.
 OX NCBI_TaxID=5722;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20382201; PubMed=10928457;
 RA Bricheux G., Coffe G., Bayle D., Brugerolle G.;
 RT "Characterization, cloning and immunolocalization of a coronin
 RT homologue in Trichomonas vaginalis";
 RL Eur. J. Cell Biol. 79:413-422(2000).
 CC -1- SIMILARITY: CONTAINS 2 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL; AJ271374; CAB76208.1; -.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 3.
 DR ProDom; PD000018; WD40; 1.
 DR SMART; SM00320; WD40; 3.
 DR PROSITE; PS50082; WD_REPEATS_2; 2.
 DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
 DR Repeat; WD repeat.
 KW REPEAT: WD repeat.
 SQ SEQUENCE 435 AA; 48204 MW; FG3ADC16D5DCE6C5 CRC64;
 Query Match 0.9%; Score 9; DB 5; Length 435;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 267 LKEENEKLR 275
 |||
 DB 412 LKEENEKLR 420
 RESULT 4
 Q9FRY1 PRELIMINARY; PRT; 512 AA.
 AC Q9FRY1;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE CRKNOX1.
 GN CRKNOX1.
 OS Ceratopteris richardii.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Filicophyta; Filicopsida; Filicales; Pteridaceae; Ceratopteris.
 OX NCBI_TaxID=49495;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HNN; TISSUE=SHOOT APEX;
 RA Sano R., Juarez C., Ito M., Banks J.A., Hasebe M.;
 RT "KNOX class of homeobox genes potentially have similar function in
 RT both sporophytic unicellular and multicellular meristems, but not in
 RT gametophytic meristems";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB043954; BAB18582.1; -.
 DR HSSP; P41778; 1DU6.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; UNKNOWN_1.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 SQ SEQUENCE 512 AA; 57195 MW; 3EBFCE074AAE080B CRC64;
 Query Match 0.9%; Score 9; DB 10; Length 512;
 Best Local Similarity 100.0%; Pred. No. 4.7;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 470 SFRSSSPAP 478

|||||

Db 255 SFRSSSPAP 263

RESULT 5

Q9HCQ3

ID Q9HCQ3 PRELIMINARY; PRT; 759 AA.

AC Q9HCQ3;

DT 01-MAR-2001 (TReMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)

DE 01-MAR-2001 (TReMBLrel. 16, Last annotation update)

DE Sperm antigen.

GN HCMOGT-1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Komori S.;

RT "HCMOGT-1";

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB041533; BAB16440.1; -.

SQ SEQUENCE 759 AA; 84920 MW; A88E5853B3E4269A CRC64;

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 266 RLKEENEKL 274

|||||

Db 480 RLKEENEKL 488

RESULT 6

Q95U22

ID Q95U22 PRELIMINARY; PRT; 949 AA.

AC Q95U22;

DT 01-DEC-2001 (TReMBLrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE GH13170p.

GN PI3K59F OR CG5373.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,

RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nucio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY058362; AAL13591.1; -.

DR FLYBase; FBgn0015277; PI3K59F.

DR InterPro; IPR001263; PI3Ka.

DR InterPro; IPR002420; PI3K_C2.

DR Pfam; PF00613; PI3Ka; 1.

DR Pfam; PF00792; PI3K_C2; 1.

DR Pfam; PF00454; PI3_Pi4_kinase; 1.

DR PROSITE; PS00915; PI3_4_KINASE_1; UNKNOWN_1.

DR PROSITE; PS00916; PI3_4_KINASE_2; UNKNOWN_1.

DR PROSITE; PS50290; PI3_4_KINASE_3; 1.

SQ SEQUENCE 949 AA; 107888 MW; 354B96CC09A3F3A2 CRC64;

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 456 LSDLSATSS 464

|||||

Db 443 LSDLSATSS 451

RESULT 7

P91635

ID P91635 PRELIMINARY; PRT; 949 AA.

AC P91635;

DT 01-MAY-1997 (TReMBLrel. 03, Created)

DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)

DE 1-phosphatidylinositol 3-kinase (EC 2.7.1.137) (PI3-kinase) (PtdIns-3-

DE kinase).

GN PI3K59F OR PI3K-59F OR CG5373.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OREGON R;

RX MEDLINE-97184674; PubMed-9032475;

RT Linossier C., Macdougall L.K., Domin J., Waterfield M.D.;

RT "Molecular cloning and biochemical characterization of a Drosophila

RT phosphatidylinositol-specific phosphoinositide 3-kinase.";

RL Biochem. J. 321:849-856(1997).

CC -!- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-ID-MYO-INOSITOL -> ADP +

CC 1-PHOSPHATIDYL-ID-MYO-INOSITOL 3-PHOSPHATE.

DR EMBL; X99912; CAA68185.1; -.

DR FLYBase; FBgn0015277; PI3K59F.

DR InterPro; IPR001263; PI3Ka.

DR InterPro; IPR002420; PI3K_C2.

DR InterPro; IPR000403; PI3_Pi4_kinase.

DR Pfam; PF00613; PI3Ka; 1.

DR Pfam; PF00792; PI3K_C2; 1.

DR Pfam; PF00454; PI3_Pi4_kinase; 1.

DR SMART; SM00145; PI3Ka; 1.

DR SMART; SM00146; PI3Kc; 1.

DR SMART; SM00142; PI3K_C2; 1.

DR PROSITE; PS00915; PI3_4_KINASE_1; 1.

DR PROSITE; PS00916; PI3_4_KINASE_2; 1.

DR PROSITE; PS50290; PI3_4_KINASE_3; 1.

DR TRANSFERASE.

SQ SEQUENCE 949 AA; 107878 MW; 4867E3ECD1E22EEC CRC64;

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 456 LSDLSATSS 464

|||||

Db 443 LSDLSATSS 451

RESULT 8

Q8WU84

ID Q8WU84 PRELIMINARY; PRT; 1020 AA.

AC Q8WU84;

DT 01-MAR-2002 (TReMBLrel. 20, Created)

DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE Hypothetical 113.9 kDa protein (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

```
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC021123; AAH21123.1; -.
DR InterPro; IPR001715; Calponin-like.
DR Pfam; PF00307; CH; 1.
DR SMART; SM00033; CH; 1.
DR PROSITE; PS0021; CH; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1020 AA; 113890 MW; 1F1D124CDF6AC8A3 CRC64;

Query Match 0.9%; Score 9; DB 4; Length 1020;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 RLKEENEKL 274
Db 463 RLKEENEKL 471
|||||

RESULT 9
Q9WLM7 PRELIMINARY; PRT; 1157 AA.
AC Q9WLM7;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE P13K59F protein.
GN P13K59F OR CG5373.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Fosler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpén G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jamal B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
```

```
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003461; AAF47030.1; -.
DR FlyBase; FBgn0015277; P13K59F.
DR InterPro; IPR000577; FGGY_kin.
DR InterPro; IPR001211; PhospholipaseA2.
DR InterPro; IPR001263; P13Ka.
DR InterPro; IPR002420; P13K_C2.
DR InterPro; IPR000403; P13_P14_kinase.
DR Pfam; PF00613; P13Ka; 1.
DR Pfam; PF00792; P13K_C2; 1.
DR Pfam; PF00454; P13_P14_kinase; 1.
DR SMART; SM00145; P13Ka; 1.
DR SMART; SM00146; P13Kc; 1.
DR SMART; SM00142; P13K_C2; 1.
DR PROSITE; PS00445; FGGY_KINASES_2; UNKNOWN_1.
DR PROSITE; PS00118; PA2_HIS; UNKNOWN_1.
DR PROSITE; PS00915; P13_4_KINASE_1; 1.
DR PROSITE; PS00916; P13_4_KINASE_2; 1.
DR PROSITE; PS0290; P13_4_KINASE_3; 1.
SQ SEQUENCE 1157 AA; 131625 MW; 54C4407B37D28B6B CRC64;

Query Match 0.9%; Score 9; DB 5; Length 1157;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 LSDLSATSS 464
Db 443 LSDLSATSS 451
|||||

RESULT 10
Q8TES3 PRELIMINARY; PRT; 1171 AA.
AC Q8TES3;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE FLJ00120 protein (Fragment).
GN FLJ00120.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA Ohara O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074049; BAB84875.1; -.
FT NON_TER 1
SQ SEQUENCE 1171 AA; 134966 MW; FA567ABBC8A703FF CRC64;

Query Match 0.9%; Score 9; DB 4; Length 1171;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 674 TSGDSFYIR 682
Db 788 TSGDSFYIR 796
|||||

RESULT 11
Q9H604 PRELIMINARY; PRT; 143 AA.
ID Q9H604
AC Q9H604;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
```

```

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CDNA: FLJ22717 fls, clone HS113737.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Oktani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK026370; BAB15463.1; -.
SQ SEQUENCE 143 AA; 15296 MW; C1C8CF31219EBDF9 CRC64;

Query Match 0.8%; Score 8; DB 4; Length 143;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 SSPAPPSQ 481
DB 11 SSPAPPSQ 18
|||||

RESULT 12
Q8THL4 PRELIMINARY; PRT; 157 AA.
AC Q8THL4;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Predicted protein.
GN MA4500
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932338;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Dearellano K., Johnson R.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talanas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE011170; AAM07840.1; -.
KW Complete proteome.
SQ SEQUENCE 157 AA; 18084 MW; A4ADC6D2CBA197AF CRC64;

Query Match 0.8%; Score 8; DB 17; Length 157;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 637 EEAVGLLR 644
DB 106 EEAVGLLR 113
|||||

RESULT 13
Q99UQ3 PRELIMINARY; PRT; 162 AA.
ID Q99UQ3

```

```

AC Q99UQ3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein SAV1215.
GN SAV1215 OR SA1058.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003361; BAB57377.1; -.
DR EMBL; AP003133; BAB42310.1; -.
DR InterPro: IPR000181; Pep_deformylase.
DR Pfam: PF01327; Pep_deformylase; 1.
DR PRINTS; PK01576; PDEFORMLASE.
DR ProDom; PD003844; Pep_deformylase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 162 AA; 18102 MW; 8E22824A4B2945AF CRC64;

Query Match 0.8%; Score 8; DB 16; Length 162;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 662 KRLLQDLE 669
DB 29 KRLLQDLE 36
|||||

RESULT 14
Q9AN71 PRELIMINARY; PRT; 223 AA.
ID Q9AN71
AC Q9AN71;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DE ID481.
DE ID481.
GN ID481.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=110SPC4;
RX MEDLINE=21101824; PubMed=11157954;
RA Gottfert M., Rothlisberger S., Kundig C., Beck C., Marty R.,
RA Hennecke H.;
RT "Potential symbiosis-specific genes uncovered by sequencing a 410-kb
RT DNA region of the Bradyrhizobium japonicum chromosome.";
RL J. Bacteriol. 183:1405-1412(2001).
DR EMBL; AF322013; AAG60905.1; -.
SQ SEQUENCE 223 AA; 25398 MW; 6F86C3C1D5D70315 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 LDLLKTRG 81

```

```

Db      29  LDLLKTRG 36
|||||
RESULT 15
Q9FSR5
ID      Q9FSR5      PRELIMINARY;      PRT;      262 AA.
AC      Q9FSR5;
DT      01-MAR-2001 (TRENBLrel. 16, Created)
DT      01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT      01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE      Putative homeodomain transcription factor.
GN      H0212B02.16.
OS      Oryza sativa (Rice).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC      Ehrhartoideae; Oryzeae; Oryza.
OX      NCBI_TaxID=4530;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Han B., Zhou B., Feng Q., Huang Y.C., Chen Z.H., Li Y., Zhu J.J.,
RA      Tang Y.S., Zhao Q., Liu Y.L., Mu J., Yu Z., Fan D.L., Chen L.,
RA      Weng Q.J., Zhang L., Lu Y.Q., Yu S.L., Zhu J., Liu X.H., Hu X.,
RA      Lei H.Y., Zhang Y.J., Wang R., Li C., Lu Y., Chen X.C., Zhang Y.,
RA      Hu H., Jia P.X., Li T., Qian Y.M., Ying K., Hong G.F.;
RT      "Oryza sativa indica (Guangluai4) genomic DNA, chromosome4, BAC
RT      clone:H0212B02."
RL      Submitted (SEP-2000) to the EMBL/GenBank/DBDJ databases.
DR      EMBL; AL442007; CAC09359.1; -.
DR      InterPro; IPR001356; Homeobox.
DR      ProDom; PD000010; Homeobox; 1.
DR      SMART; SM00389; HOX; 1.
DR      PROSITE; PS0071; HOMEBOX.2; 1.
KW      DNA-binding; Homeobox; Nuclear protein.
SQ      SEQUENCE 262 AA; 29004 MW; C54C573691F8CBF8 CRC64;

Query Match      0.8%; Score 8; DB 10; Length 262;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 SSSPAPPS 480
Db 19 SSSPAPPS 26
|||||

```

Search completed: January 22, 2003, 08:59:39
Job time : 54 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 20.8535 Seconds
(without alignments)
683.714 Million cell updates/sec

Title: US-09-767-215-2_COPY_10_116

Perfect score: 560

Sequence: 1 ALFALDETLWEMESHRR.....PDVYTLVTGLQPDVDFSNFS 107

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	560	100.0	1004	22 AAE07164	Human caspase recr
2	560	100.0	1139	22 AAE07165	Human predicted ca
3	270	48.2	1147	22 AAU01207	Human caspase recr
4	246	43.9	1032	22 AAU01206	Human caspase recr
5	246	43.9	1032	22 AAU073247	Human plakoglobin
6	217	38.8	536	22 AAU01204	Rat caspase recrui
7	209	37.3	174	21 ABA41067	Human ORFX ORF831
8	209	37.3	174	23 ABP10782	Human ORFX protein
9	209	37.3	366	23 ABA95617	Human protein sequ
10	209	37.3	536	22 AAU01205	Human caspase recr

11	100	17.9	237	23	AAU73245	Human plakoglobin
12	81	14.5	233	21	AAU59412	Human CIPER protei
13	78	13.9	233	21	AAU59413	Mouse CIPER protei
14	70.5	12.6	845	22	ABG62651	Drosophila melanog
15	70.5	12.6	1266	23	AAE22544	CTAL-OVA-DD fusion
16	69	12.3	519	21	AAU44570	Human voltage-gate
17	69	12.3	1416	22	AAU30862	Novel human secret
18	68	12.1	385	22	AAU93386	Human polypeptide,
19	68	12.1	412	22	AAU93214	Human polypeptide,
20	68	12.1	519	21	AAU44565	Human voltage-gate
21	68	12.1	519	21	AAU44569	Human voltage-gate
22	68	12.1	519	21	AAU44571	Human voltage-gate
23	68	12.1	519	21	AAU44572	Human voltage-gate
24	67.5	12.1	763	22	AAU90700	C glutamic prote
25	67.5	12.1	2443	22	ABG60521	Drosophila melanog
26	66.5	11.9	497	17	AAU81548	Hemolysin. Serpul
27	66	11.8	477	21	ABA42919	Human ORFX ORF2683
28	65.5	11.7	209	23	AAU10658	Mouse LICAM Fn4-5
29	65.5	11.7	309	23	AAU10657	Mouse LICAM Fn3-5
30	65.5	11.7	397	23	AAU10656	Mouse LICAM F80 ex
31	65.5	11.7	409	19	AAU70604	Ankyrin protein fr
32	65.5	11.7	409	19	AAU76772	D. immitis ankyrin
33	65.5	11.7	409	21	ABA11585	D. immitis ankyrin
34	65.5	11.7	409	23	AAO21364	Ankyrin protein se
35	65.5	11.7	422	21	ABA11617	D. immitis ankyrin
36	65.5	11.7	422	23	AAO21396	Ankyrin protein se
37	65.5	11.7	447	23	AAU10670	Mouse LICAM Fn4-5
38	65.5	11.7	448	22	ABA52467	Mycobacterium tube
39	65.5	11.7	547	23	AAU10669	Mouse LICAM Fn3-5
40	65.5	11.7	635	23	AAU10668	Mouse LICAM F80 fu
41	65.5	11.7	667	23	ABB77383	Wheat acyl CoA oxi
42	65.5	11.7	669	23	ABB77378	Rice acyl CoA oxi
43	65.5	11.7	1260	23	AAU10649	Mouse LICAM protei
44	65.5	11.7	1745	19	AAU70608	Full length ankyri
45	65.5	11.7	1745	19	AAU76776	D. immitis ankyrin

ALIGNMENTS

RESULT 1
AAE07164
ID AAE07164 standard; Protein; 1004 AA.
XX
AC AAE07164;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human caspase recruitment domain-14 (CARD-14).
XX
KW Human; caspase recruitment domain-14; CARD-14; chromosome 17;
KW nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy;
KW autoimmune disorder; systemic lupus erythematosus; neurological disorder;
KW Alzheimer's disease; Parkinson's disease; inflammatory disorder; anaemia;
KW haematological disorder; myelodysplastic syndrome; myocardial infarction;
KW stroke; immune disorder; Crohn's disease; allergic rhinitis; infection;
KW cell signalling disorder; cytostatic; immunosuppressive; neurotropic;
KW neuroprotective; antiviral; antibacterial.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 6..9
FT /note= "CAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Domain 10..116
FT /label= CARD_domain
FT Modified-site 12..15
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 18..21
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 25..27
FT /note= "Protein kinase C phosphorylation site"

FT	Modified-site	60..62	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	/note= "N-glycosylation site"
FT	Modified-site	91..93	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	602..605
FT	Modified-site	114..117	/note= "N-glycosylation site"	FT	Modified-site	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	117..122	/note= "N-myristoylation site"	FT	Modified-site	634..637
FT	Modified-site	121..123	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	/note= "Casein kinase II phosphorylation site"
FT	Domain	126..420	/label= Coiled_Coil_domain	FT	Modified-site	653..655
FT	Modified-site	130..135	/note= "N-myristoylation site"	FT	Modified-site	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	134..137	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	674..677
FT	Modified-site	161..166	/note= "N-myristoylation site"	FT	Modified-site	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	165..168	/note= "N-myristoylation site"	FT	Domain	676..745
FT	Modified-site	220..227	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	/label= SH3_domain
FT	Modified-site	221..224	/note= "Tyrosine kinase phosphorylation site"	FT	Modified-site	714..719
FT	Domain	239..325	/label= "Casein kinase II phosphorylation site"	FT	Modified-site	/note= "N-myristoylation site"
FT	Modified-site	240..243	/label= k-Box_domain	FT	Modified-site	725..727
FT	Modified-site	250..252	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	253..256	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	725..728
FT	Modified-site	259..262	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	280..293	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	733..738
FT	Modified-site	290..293	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	297..300	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	/note= "N-myristoylation site"
FT	Modified-site	307..309	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	737..740
FT	Modified-site	307..310	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	/note= "N-glycosylation site"
FT	Modified-site	359..365	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	366..368	/note= "Tyrosine kinase phosphorylation site"	FT	Peptide	/note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT	Modified-site	366..369	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	785..793
FT	Modified-site	378..381	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	/note= "Peroxisomal targeting signal"
FT	Modified-site	384..386	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	796..799
FT	Region	385..406	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	449..452	/note= "Leucine zipper pattern"	FT	Modified-site	800..805
FT	Modified-site	463..466	/note= "Casein kinase II phosphorylation site"	FT	Domain	826..1004
FT	Modified-site	463..465	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	/label= Guanylate_kinase_domain
FT	Modified-site	470..472	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	842..844
FT	Modified-site	501..504	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	511..516	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	860..863
FT	Domain	568..660	/label= PDZ_domain	FT	Modified-site	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	587..592	/note= "N-myristoylation site"	FT	Modified-site	868..870
FT	Modified-site	589..592	/note= "N-myristoylation site"	FT	Region	/note= "Protein kinase C phosphorylation site"
				FT	Modified-site	870..872
				FT	Modified-site	/note= "RGD cell attachment sequence"
				FT	Modified-site	893..896
				FT	Modified-site	/note= "Casein kinase II phosphorylation site"
				FT	Modified-site	926..929
				FT	Peptide	/note= "Casein kinase II phosphorylation site"
				FT	Modified-site	941..949
				FT	Modified-site	/note= "Peroxisomal targeting signal"
				FT	Modified-site	944..947
				FT	Modified-site	/note= "Casein kinase II phosphorylation site"
				FT	Modified-site	976..979
				FT	Modified-site	/note= "Casein kinase II phosphorylation site"
				FT	Modified-site	980..985
				FT	Modified-site	/note= "N-myristoylation site"
				FT	Modified-site	1002..1004
				FT	Modified-site	/note= "Protein kinase C phosphorylation site"
				XX		W0200159065-A2.
				XX		16-AUG-2001.
				PD		22-JAN-2001; 2001WO-US02087.
				XX		09-FEB-2000; 2000US-0181159.
				XX		(MILL-) MILLENNIUM PHARM INC.
				XX		Bertin J;
				XX		WPI; 2001-497073/54.
				DR		N-PSDB; RAD13447.
				XX		An isolated caspase recruitment domain polypeptide useful for

CC mapping, tissue typing), predictive medicine (prognostic assays,
CC monitoring clinical trials, and therapy (treatment and prophylaxis)). The
CC CARD polypeptide may be used to screen for drugs that bind to and/or
CC modulate it. CARD sequences are potential targets for regulating
CC inflammation, cancer, NF-kappaB signalling, stress-related response and
CC apoptosis in human disease A host cell containing a polynucleotide
CC encoding CARD can be used to create transgenic animals.

XX Sequence 1147 AA;

Query Match 48.2%; Score 270; DB 22; Length 1147;
Best Local Similarity 51.5%; Pred No. 4.3e-26;
Matches 51; Conservative 22; Mismatches 26; Indels 0; Gaps 0;

QY 6 DEPTLWMESHRRIVRCPSRLPYLRQAKVLQCLDEEVLHSPRLTNSAMRAGHLL 65

:|: ||| :|: | :|: ||||| ||| :|: ||||| :|: | ||| ||

Db 11 EEDALWENECNRHMLSLRINPAKLPYLROCKVIDEQDEDEVLNAPMLPSKINRAGRL 70

:|: ||| :|: | :|: ||||| :|: | ||||| :|: | ||| ||

QY 66 DLLKTRKNGATAFLSLKFNPDVYTLVTGLQPDVDFS 104

:|: ||| :|: | :|: ||||| :|: | ||||| :|: | ||| ||

Db 71 DILHTGQGVVVFLESLEFYYPYELVKLVTKGKPTRRFS 109

RESULT 4

AAU01206 ID AAU01206 standard; Protein; 1032 AA.

XX AC AAU01206;

DT 12-SEP-2001 (first entry)

XX DE Human caspase recruitment domain, CARD-10 polypeptide.

XX KW Human; caspase recruitment domain; CARD-10; Bcl-10; NF-kappaB;

XX KW apoptosis; hyperproliferative disorder; autoimmune; neurological;

XX KW inflammatory disorder; viral infection; stress-related response.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Modified-site 15..20 /note= "N-myristoylation site"

FT Modified-site 19..21 /note= "Casein kinase II phosphorylation site"

FT Domain 23..123 /note= "CARD domain"

FT Modified-site 68..70 /note= "Protein kinase C phosphorylation site"

FT Modified-site 76..79 /note= "N-glycosylation site"

FT Modified-site 78..80 /note= "Protein kinase C phosphorylation site"

FT Modified-site 88..91 /note= "Amidation site"

FT Modified-site 112..115 /note= "Casein kinase II phosphorylation site"

FT Modified-site 113..118 /note= "N-myristoylation site"

FT Domain 147..457 /note= "Coiled coil domain"

FT Modified-site 201..207 /note= "Tyrosine kinase phosphorylation site"

FT Region 230..251 /note= "Leucine zipper homology region"

FT Modified-site 242..245 /note= "Casein kinase II phosphorylation site"

FT Modified-site 293..295 /note= "Protein kinase C phosphorylation site"

FT Modified-site 293..296 /note= "Casein kinase II phosphorylation site"

FT Modified-site 309..314 /note= "N-myristoylation site"

FT Modified-site 313..315

FT Modified-site /note= "Protein kinase C phosphorylation site"

FT Domain /note= "Casein kinase II phosphorylation site"

FT Modified-site /note= "Tropomyosin domain"

FT Region /note= "Casein kinase II phosphorylation site"

FT Modified-site /note= "Leucine zipper homology region"

FT Domain /note= "Casein kinase II phosphorylation site"

FT Modified-site /note= "MAGUK domain"

FT Modified-site /note= "N-glycosylation site"

FT Modified-site /note= "Casein kinase II phosphorylation site"

FT Modified-site /note= "N-myristoylation site"

FT Modified-site /note= "Protein kinase C phosphorylation site"

FT Modified-site /note= "Casein kinase II phosphorylation site"

FT Modified-site /note= "Protein kinase C phosphorylation site"

FT Modified-site /note= "N-myristoylation site"

FT Modified-site /note= "Casein kinase II phosphorylation site"

FT Modified-site /note= "N-glycosylation site"

FT Modified-site /note= "Protein kinase C phosphorylation site"

FT Modified-site /note= "Glycosaminoglycan attachment site"

FT Modified-site /note= "Protein kinase C phosphorylation site"

FT Modified-site /note= "N-myristoylation site"

FT Modified-site /note= "Casein kinase II phosphorylation site"

FT Modified-site /note= "N-glycosylation site"

FT Modified-site /note= "Protein kinase C phosphorylation site"

FT Modified-site /note= "Protein kinase C phosphorylation site"

FT Domain /note= "SH3 domain"

FT Modified-site /note= "N-glycosylation site"

FT Modified-site /note= "Casein kinase II phosphorylation site"

FT Modified-site /note= "Tyrosine kinase phosphorylation site"

FT Modified-site /note= "Casein kinase II phosphorylation site"

FT Modified-site /note= "Protein kinase C phosphorylation site"

FT Modified-site /note= "Casein kinase II phosphorylation site"

FT Modified-site /note= "N-myristoylation site"

FT Modified-site /note= "Protein kinase C phosphorylation site"

FT Domain /note= "N-myristoylation site"

FT Modified-site /note= "Guanylate kinase (GUK) domain"

FT Modified-site /note= "Protein kinase C phosphorylation site"

FT Modified-site /note= "Protein kinase C phosphorylation site"

FT Modified-site /note= "Casein kinase II phosphorylation site"


```

RESULT 7
AAB41067
ID AAB41067 standard; Protein; 174 AA.
XX AC AAB41067;
XX DT
XX 08-FEB-2001 (first entry)
XX DE Human ORFX ORF831 polypeptide sequence SEQ ID NO:1662.
XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX OS Homo sapiens.
XX PN WO200058473-A2.
XX PD
XX 05-OCT-2000.
XX PF 31-MAR-2000; 2000WO-US08621.
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX PA (CURA-) CURAGEN CORP.
XX PI Shinkets RA, Leach M;
XX WPI; 2000-602362/57.
XX N-PSDB; AAC75276.
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX Claim 11; Page 1326; 5507pp; English.
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antithyroid; and antianemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.

```

```

XX SQ Sequence 174 AA:
Query Match 37.3%; Score 209; DB 21; Length 174;
Best Local Similarity 44.4%; Pred. No. 3.5e-19;
Matches 44; Conservative 16; Mismatches 39; Indels 0; Gaps 0;
Qy 6 DEETLWEMESHRRIVRCICPSRLTPYLQAKVLCQDEEVLHSPRLTNSAMRAGHLL 65
Db 71 NDDECNVLEGFRVTLTSDIDPSRITPYLRQCKVLNPDDEQVLSDPNLVIRKRVGVLL 130
Qy 66 DLLKTRGRNGAIFLESLEKFNPDVYTLVTGLQPDVDFS 104
Db 131 DILQRTGHKGVAFLSELYLPOLYKVKVTKGEPARVFS 169
RESULT 8
ABP10782
ID ABP10782 standard; Protein; 174 AA.
XX AC ABP10782;
XX DT 24-JUN-2002 (first entry)
XX DE Human ORFX protein sequence SEQ ID NO:21546.
XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX OS Homo sapiens.
XX PN WO200192523-A2.
XX PD 06-DEC-2001.
XX PF 29-MAY-2001; 2001WO-US10836.
XX PR 30-MAY-2000; 2000US-206132P.
XX PR 29-AUG-2000; 2000US-228716P.
XX PA (CURA-) CURAGEN CORP.
XX PI Shinkets RA, Leach MD;
XX WPI; 2002-106308/14.
XX N-PSDB; ABN26534.
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and autoimmune disorders -
XX Disclosure; SEQ ID 21546; 1037pp; English.
XX The present invention describes substantially purified human proteins
XX (referred to as open reading frame, ORFX, where X is 1-11491 (see table 1
XX in the specification). ABN15762 to ABN27252 encode the human ORFX
XX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
XX treating or preventing a pathology associated with an ORFX-associated
XX disorder in humans, and in the manufacture of a medicament for treating a
XX syndrome associated with ORFX-associated disorder. ORFX polynucleotide
XX sequences can be used in gene therapy. ORFX sequences can be used in the
XX treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
XX psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
XX osteoarthritis, neurodegenerative disorders, disorders related to organ
XX transplantation, cardiovascular diseases, diabetes mellitus, systemic
XX lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
XX storage disease, various immune deficiencies and disorders, infectious

```

CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 174 AA;
SQ
Query Match 37.3%; Score 209; DB 23; Length 174;
Best Local Similarity 44.4%; Pred. No. 3.5e-19;
Matches 44; Conservative 16; Mismatches 39; Indels 0; Gaps 0;
QY 6 DEETLWEMESHRRIVRCICPSRLTPYLROAKVLCQDDEEVHLSPRLTNSAMRAGHLL 65
Db 71 NDDECNVLEGFRTVTSVIDPSRITPYLRQCKVLNPDDEEQVLSDPNLVIRKRVGVLL 130
QY 66 DLLKTRGKNCATAFLESKLFHNPDVVTLVTGLQPDVDES 104
Db 131 DILQRTGHGYVAFLESLELYYPOLYKVTGTGEPARVFS 169
RESULT 9
AAB95617
ID AAB95617 standard; Protein; 366 AA.
XX
AC AAB95617;
XX
XX 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:18328.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 18328; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 366 AA;
SQ
Query Match 37.3%; Score 209; DB 22; Length 366;
Best Local Similarity 44.4%; Pred. No. 1e-18;
Matches 44; Conservative 16; Mismatches 39; Indels 0; Gaps 0;
QY 6 DEETLWEMESHRRIVRCICPSRLTPYLROAKVLCQDDEEVHLSPRLTNSAMRAGHLL 65
Db 6 NDDECNVLEGFRTVTSVIDPSRITPYLRQCKVLNPDDEEQVLSDPNLVIRKRVGVLL 65
QY 66 DLLKTRGKNCATAFLESKLFHNPDVVTLVTGLQPDVDES 104
Db 66 DILQRTGHGYVAFLESLELYYPOLYKVTGTGEPARVFS 104
RESULT 10
AAU01205
ID AAU01205 standard; Protein; 536 AA.
XX
AC AAU01205;
XX
XX 12-SEP-2001 (first entry)
XX
XX Human caspase recruitment domain, CARD-9 polypeptide.
XX
XX Human; caspase recruitment domain; CARD-9; Bcl-10; NF-kappaB;
XX apoptosis; hyperproliferative disorder; autoimmune; neurological;
XX inflammatory disorder; viral infection; stress-related response.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Modified-site 2..5
FT /note= "Casein kinase II phosphorylation site"
FT Domain 7..98
FT /note= "CARD domain"
FT Modified-site 23..26
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 92..95
FT /note= "cAMP- and cGMP-dependent protein kinase
FT phosphorylation site"
FT Modified-site 95..97
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 95..98
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 138..140
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 138..141
FT /note= "Casein kinase II phosphorylation site"
FT Domain 140..416
FT /note= "Coiled coil domain"
FT Modified-site 176..183
FT /note= "Tyrosine kinase phosphorylation site"
FT Region 197..213
FT /note= "Indole-3-glycerol phosphate synthase
FT homology region"

FT Modified-site 228..231
FT /note= "cAMP- and cGMP-dependent protein kinase
FT phosphorylation site"
FT
FT Modified-site 231..233
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 267..270
FT /note= "Casein kinase II phosphorylation site"
FT Region 285..338
FT /note= "cysteine rich repeat homology region"
FT Modified-site 303..305
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 333..336
FT /note= "cAMP- and cGMP-dependent protein kinase
FT phosphorylation site"
FT
FT Modified-site 363..366
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 425..428
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 431..433
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 450..452
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 453..458
FT /note= "N-myristoylation site"
FT Modified-site 460..462
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 481..486
FT /note= "N-myristoylation site"
FT Modified-site 483..486
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 524..527
FT /note= "N-glycosylation site"
FT Modified-site 526..529
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 527..532
FT /note= "N-myristoylation site"
FT Modified-site 531..534
FT /note= "Casein kinase II phosphorylation site"
FT
FT WO200104068-A2.
FT
FT 07-JUN-2001.
FT
FT 01-DEC-2000; 2000WO-US32716.
FT
FT 03-DEC-1999; 99US-0168780.
FT 18-FEB-2000; 2000US-0507533.
FT 23-FEB-2000; 2000US-0513904.
FT 10-OCT-2000; 2000US-0685791.
FT
FT (MILL-) MILLENNIUM PHARM INC.
FT
FT Bertin J;
FT
FT WPI: 2001-367809/38.
FT N-PSDB; AAS05387.
FT
FT Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10,
FT CARD-11, useful as targets for therapy, as immunogens, and in screening
FT and detection assays -
FT
FT Claim 9; Fig 5A-5B; 145pp; English.
FT
FT The present sequence represents novel human caspase recruitment
FT domain, CARD-9. The polynucleotide encoding this sequence was isolated
FT from a human megakaryocyte cDNA library. Also described are novel human
FT sequences for CARD-10 and CARD-11 (AAU01206, AAU01207) and rat CARD-9
FT (AAU01204). CARD-9, CARD-10 and CARD-11 interact with Bcl-10 which is
FT thought to activate nuclear factor (NF)-kappaB and apoptosis. The
FT sequences of the invention can be used for treating a disorder
FT associated with abnormal levels of apoptosis by modulating the
FT expression or activity of CARD-9, CARD-10, or CARD-11. They can
FT be used for the treatment of hyperproliferative disorders (e.g.

CC cancer), autoimmune disorders (e.g. systemic lupus erythematosus),
CC neurological disorders (e.g. Alzheimer's disease), inflammatory disorders
CC (e.g. Crohn's disease), and viral infection (e.g. HIV). The CARD
CC polypeptide, polynucleotide and an antibody which selectively binds to
CC CARD can be used in screening and detection assays (e.g. chromosomal
CC mapping, tissue typing), predictive medicine (prognostic assays,
CC monitoring clinical trials, and therapy (treatment and prophylaxis). The
CC CARD polypeptide may be used to screen for drugs that bind to and/or
CC modulate it. CARD sequences are potential targets for regulating
CC inflammation, cancer, NF-kappaB signalling, stress-related response and
CC apoptosis in human disease. A host cell containing a polynucleotide
CC encoding CARD can be used to create transgenic animals.
XX
SQ Sequence 536 AA;

Query Match 37.3%; Score 209; DB 22; Length 536;
Best Local Similarity 44.4%; Pred. No. 1.7e-18;
Matches 44; Conservative 16; Mismatches 39; Indels 0; Gaps 0;

QY 6 DEETLWEMESHRRHIVRCICPSRLTPYLROAKVLCQDDEEVLHSPRLTNSMRAGHLL 65
DB 6 NDDECMNVLEGFVRVLTSTVIDPSRITPYLRCKVLNPDDEQVLSDPNLVIRKRVGVLL 65

QY 66 DLLKTRGNGAIAFLESILKFHNPVYTLVTGLQPDVDFS 104
DB 66 DILQRTGHGYVAFLESLLEYYPOLYKVKTKGEPARVFS 104

RESULT 11
AAU73245
ID AAU73245 standard; Protein; 237 AA.
XX
AC AAU73245;
XX
DT 12-MAR-2002 (first entry)
XX
DE Human plakoglobin interacting protein #1.
XX
KW Human; plakoglobin; cytostatic; osteopathic; dermatological; cardiant;
KW plakoglobin related disease; skin carcinoma; acantholytic disease;
KW basal cell carcinoma; squamous cell carcinoma; Naxos disease;
KW extramammary Paget's disease; heart disease; skin blistering;
KW subcorneal acantholysis; Grover's disease; Hailey-Hailey's disease;
KW Darier's disease; ectodermal dysplasia; skin fragility syndrome.
XX
OS Homo sapiens.
XX
PN WO200185933-A2.
XX
PD 15-NOV-2001.
XX
PF 02-MAY-2001; 2001WO-EP04872.
XX
PR 09-MAY-2000; 2000EP-0201668.
XX
PA (VLAA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX
PI Van Roy F, Bonne S, Vanlandschoot A;
XX
DR WPI: 2002-062246/08.
DR N-PSDB; AAS98201.
XX
PT New polypeptide, useful for treating skin carcinoma or acantholytic
PT disease such as Grover's and Darier's disease, comprises a protein
PT interacting with human plakoglobin and involved in transduction of
PT plakoglobin related signal to nucleus -
XX
PS Claim 2; Page 42-43; 98pp; English.
XX
CC The invention relates to an isolated plakoglobin interacting polypeptide
CC (I). (I) is useful as a medicament and in the manufacture of a
CC medicament for treating plakoglobin related diseases, such as skin
CC carcinoma or an acantholytic disease, and to screen compounds that

CC interfere with the interaction of the polypeptide with plakoglobin
 CC The plakoglobin related diseases include basal cell carcinoma, squamous
 CC cell carcinoma, extramammary Paget's disease, Naxos disease, heart
 CC diseases, skin blistering and acantholytic diseases such as subcorneal
 CC acantholysis, Grover's disease, Hailey-Hailey's disease or Darier's
 CC disease, and ectodermal dysplasia/skin fragility syndrome. AAU73245-
 CC AAU73254 represent novel human plakoglobin interacting protein
 CC amino acid sequences and related sequences of the invention.
 XX
 SQ Sequence 237 AA;
 Query Match 17.9%; Score 100; DB 23; Length 237;
 Best Local Similarity 51.4%; Pred. No. 0.00014;
 Matches 18; Conservative 10; Mismatches 7; Indels 0; Gaps 0;
 QY 65 DLLKTRKNGAIAFLESILKFNPDVYTLVTGLQP 99
 Db 1 MDILRCRGKRGVEAFLEALEFFYPPEHFTLLTGQEP 35
 :||: ||| | ||||:|:|: | :||:|:| :|
 :||: ||| | ||||:|:|: | :||:|:| :|
 RESULT 12
 AAY59412
 ID AAY59412 standard; Protein; 233 AA.
 XX
 AC AAY59412;
 XX
 DT 21-MAR-2000 (first entry)
 DE Human CIPER protein sequence.
 XX
 KW RICK; mouse; RIP-like interacting CLARP kinase; apoptosis regulator; ARC;
 KW caspase-8; caspase-10; CD95 signalling; apoptosis signalling pathway;
 KW CIDE-A; CIDE-B; DREP-1; diagnosis; cell growth; apoptosis dysregulation;
 KW gene therapy; cardiac disorder; cancer; neurodegenerative disorder; AIDS;
 KW aplastic anaemia; ischaemic injury; toxin-induced liver disease;
 KW CIPER.
 XX
 OS Homo sapiens.
 XX
 PN WO9955134-A2.
 XX
 PD 04-NOV-1999.
 XX
 PF 27-APR-1999; 99WO-US09183.
 XX
 PR 27-APR-1998; 98US-0069023.
 XX
 PA (UNMI) UNIV MICHIGAN.
 XX
 PI Nunez G, Inohara N, Koseki T;
 XX
 DR WPI: 2000-072163/06.
 DR N-PSDB; AAZ48767.
 XX
 PT Compositions for identifying apoptosis signalling pathway inhibitors
 useful for treating diseases -
 XX
 PS Disclosure; Fig 20b; 93pp; English.
 XX
 CC This sequence is the human CIPER protein. The invention relates to
 CC the human RICK (RIP-like interacting CLARP kinase) protein. The RICK
 CC protein acts as a positive regulator of apoptosis, potentiating apoptosis
 CC induced by caspase-8 and caspase-10 during CD95 signalling. The invention
 CC provides methods for identifying apoptosis signalling pathway inhibitors
 CC and activators, and methods and compositions for screening compounds
 CC which will modulate the interactions of the various compositions
 CC identified: ARC, RICK, and the CIDE family of activators (CIDE-A, CIDE-B
 CC and DREP-1). RICK is useful in screening assays for agents, useful in the
 CC diagnosis, prognosis or treatment of disease associated with excess cell
 CC growth and dysregulation of apoptosis. Complexes containing RICK and
 CC CLARP can be used in drug screening assays to identify inhibitor
 CC molecules blocking CD95-mediated apoptosis. Overexpression of ARC in an
 CC in vitro cell system can be used to identify inhibitors of the enzymatic

CC activity of caspase-8. Identification of ARC-like inhibitory compounds
 CC may be useful for gene therapy treatment of disease with increased cell
 CC death in muscle tissue and cardiac disorders. Therapeutic compositions of
 CC CIDEs can be used to treat e.g. cancer, AIDS, neurodegenerative
 CC disorders, aplastic anaemia, ischaemic injury, and toxin-induced liver
 CC disease. AntiRICK antibodies can be used as reagents for the preparation
 CC or affinity chromatography media, and for diagnostically measuring RICK
 CC levels. A specific inhibitor of an essential step in the biochemistry of
 CC apoptosis is needed. RICK interaction with intracellular factors such as
 CC CLARP and FADD appears to be essential for apoptosis, inhibitors of RICK
 CC binding to intracellular apoptosis factors are potential drug candidates.
 XX
 SQ Sequence 233 AA;
 Query Match 14.5%; Score 81; DB 21; Length 233;
 Best Local Similarity 36.5%; Pred. No. 0.045;
 Matches 31; Conservative 14; Mismatches 32; Indels 8; Gaps 4;
 QY 4 ALDEETLWEM----MESHRRIVRCICPSRLPYLRQAKVLCOLDEEVLHSPRLTNSAM 59
 Db 7 SLTEEDLTVKKDALENLRVYLCEKIIAERHFDHLRAKKILSREDETEI--SCR-TSSRK 63
 :||: ||| | :|: | :|: | :||:|:| :|:|:|
 :||: ||| | :|: | :|: | :||:|:| :|:|:|
 QY 60 RAGHLIDLKTRKNGATAFLESILK 84
 Db 64 RAGKLLDYLQENPK-GLDTLVESIR 87
 :||: ||| | :|: | :|: | :||:|:| :|:|:|
 :||: ||| | :|: | :|: | :||:|:| :|:|:|
 RESULT 13
 AAY59413
 ID AAY59413 standard; Protein; 233 AA.
 XX
 AC AAY59413;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Mouse CIPER protein sequence.
 XX
 KW RICK; mouse; RIP-like interacting CLARP kinase; apoptosis regulator; ARC;
 KW caspase-8; caspase-10; CD95 signalling; apoptosis signalling pathway;
 KW CIDE-A; CIDE-B; DREP-1; diagnosis; cell growth; apoptosis dysregulation;
 KW gene therapy; cardiac disorder; cancer; neurodegenerative disorder; AIDS;
 KW aplastic anaemia; ischaemic injury; toxin-induced liver disease;
 KW CIPER.
 XX
 OS Mus sp.
 XX
 PN WO9955134-A2.
 XX
 PD 04-NOV-1999.
 XX
 PF 27-APR-1999; 99WO-US09183.
 XX
 PR 27-APR-1998; 98US-0069023.
 XX
 PA (UNMI) UNIV MICHIGAN.
 XX
 PI Nunez G, Inohara N, Koseki T;
 XX
 DR WPI: 2000-072163/06.
 DR N-PSDB; AAZ48768.
 XX
 PT Compositions for identifying apoptosis signalling pathway inhibitors
 useful for treating diseases -
 XX
 PS Disclosure; Fig 20d; 93pp; English.
 XX
 CC This sequence is the mouse CIPER protein. The invention relates to
 CC the human RICK (RIP-like interacting CLARP kinase) protein. The RICK
 CC protein acts as a positive regulator of apoptosis, potentiating apoptosis
 CC induced by caspase-8 and caspase-10 during CD95 signalling. The invention
 CC provides methods for identifying apoptosis signalling pathway inhibitors
 CC and activators, and methods and compositions for screening compounds
 CC which will modulate the interactions of the various compositions

THIS PAGE BLANK (USPTO)


```
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-069-023-38

Query Match      13.9%; Score 78; DB 4; Length 233;
Best Local Similarity 35.3%; Pred. No. 0.015;
Matches 30; Conservative 14; Mismatches 33; Indels 8; Gaps 4;

QY  4 ALDEETLWEM----MESIRHVRICPSRLTPYLROAKVLCQDDEEVLSHSPRLTNSAM 59
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  7 SLTEEDLTVEVKDALENLRVLCERIIAERHFDHLRAKKILSREDTEI--SCR-TSSRK 63
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY  60 RAGHLDLLKTRGKNGATAFLESK 84
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  64 RAGKLDYDLYQ-ENRGLDTLVESIR 87
    :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT 3
US-08-252-492-2
; Sequence 2, Application US/08252492
; Patent No. 5698394
; GENERAL INFORMATION:
; APPLICANT: Duhamel, Gerald E.
; APPLICANT: Elder, Robert
; TITLE OF INVENTION: Nucleotide Sequences and Methods for
; TITLE OF INVENTION: Detection of Serpulina Hyodysenteriae
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5698394west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,492
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalchuk, Katherine M.
; REGISTRATION NUMBER: 36,848
; REFERENCE/DOCKET NUMBER: 9338.25US01
; TELEPHONE: 612-332-5300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-252-492-2

Query Match      11.9%; Score 66.5; DB 1; Length 497;
Best Local Similarity 26.1%; Pred. No. 1.8;
Matches 24; Conservative 18; Mismatches 33; Indels 17; Gaps 4;

QY  3 TALDEETLWEMESHRRIVRCICPSRLTPYLROAKVLCQDDEEVLSHSPRLTNSAMRAG 62
    :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  342 TVIDADALYLMFESTLNELNNFI---ITPHYFEKLTQINHIEVLENPYQA----- 391
    :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY  63 HLDDLKTRGK-NGATAFLESKLF--HNPVDY 91
    :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  392 ----LLIYREKTNASIVLKDAVSFLMHENDIY 419
    :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT 4
US-08-252-492-2

Query Match      11.9%; Score 66.5; DB 1; Length 497;
Best Local Similarity 26.1%; Pred. No. 1.8;
Matches 24; Conservative 18; Mismatches 33; Indels 17; Gaps 4;

QY  3 TALDEETLWEMESHRRIVRCICPSRLTPYLROAKVLCQDDEEVLSHSPRLTNSAMRAG 62
    :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  342 TVIDADALYLMFESTLNELNNFI---ITPHYFEKLTQINHIEVLENPYQA----- 391
    :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY  63 HLDDLKTRGK-NGATAFLESKLF--HNPVDY 91
    :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  392 ----LLIYREKTNASIVLKDAVSFLMHENDIY 419
    :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT 4
```

```
US-08-727-126-2
; Sequence 2, Application US/08727126
; Patent No. 5869630
; GENERAL INFORMATION:
; APPLICANT: Duhamel, Gerald E.
; APPLICANT: Elder, Robert
; TITLE OF INVENTION: Nucleotide Sequences and Methods for
; TITLE OF INVENTION: Detection of Serpulina Hyodysenteriae
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5869630west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/727,126
; FILING DATE: 08-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/252,492
; FILING DATE: 01-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalchuk, Katherine M.
; REGISTRATION NUMBER: 36,848
; REFERENCE/DOCKET NUMBER: 9338.25US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-727-126-2

Query Match      11.9%; Score 66.5; DB 2; Length 497;
Best Local Similarity 26.1%; Pred. No. 1.8;
Matches 24; Conservative 18; Mismatches 33; Indels 17; Gaps 4;

QY  3 TALDEETLWEMESHRRIVRCICPSRLTPYLROAKVLCQDDEEVLSHSPRLTNSAMRAG 62
    :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  342 TVIDADALYLMFESTLNELNNFI---ITPHYFEKLTQINHIEVLENPYQA----- 391
    :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY  63 HLDDLKTRGK-NGATAFLESKLF--HNPVDY 91
    :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  392 ----LLIYREKTNASIVLKDAVSFLMHENDIY 419
    :||:|:|:|:|:|:~|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT 5
US-08-942-761-2
; Sequence 2, Application US/08942761
; Patent No. 6068843
; GENERAL INFORMATION:
; APPLICANT: Duhamel, Gerald E.
; APPLICANT: Elder, Robert
; TITLE OF INVENTION: Nucleotide Sequences and Methods for
; TITLE OF INVENTION: Detection of Serpulina Hyodysenteriae
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 6068843west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,761
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/252,492
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalchyk, Katherine M.
; REGISTRATION NUMBER: 36,848
; REFERENCE/DOCKET NUMBER: 9338.25U501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-942-761-2

Query Match 11.9%; Score 66.5; DB 3; Length 497;
Best Local Similarity 26.1%; Pred. No. 1.8;
Matches 24; Conservative 18; Mismatches 33; Indels 17; Gaps 4;

QY 3 TALDEETLWEMESHRRIVRCIPSRTPPYLRQAKVLCQLEDEEVLSHSPRLTNSAMRAG 62
Db 342 TVIDADALYKTNELNNFI---ITPHIYEKLTQINHIEVLNPYQA----- 391

QY 63 HLLDLKTRGK-NGAIAFLESLEK--HNPDVY 91
Db 392 ----LLIYREKTNASIVLKDAVSFLMHENDIY 419

RESULT 6
US-08-341-843B-13
; Sequence 13, Application US/08341843B
; Patent No. 5872225
; GENERAL INFORMATION:
; APPLICANT: Lemmon, Vance
; TITLE OF INVENTION: A Method for Characterizing the
; TITLE OF INVENTION: Nucleotide Sequence of LICAM and
; Patent No. 5872225
; TITLE OF INVENTION: the Nucleotide Sequence
; TITLE OF INVENTION: Characterized Thereby
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; MEDIUM TYPE: storable
; COMPUTER: Compaq Prolinea 5100e
; OPERATING SYSTEM: DOS 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/341,843B
; FILING DATE: No. 5872225ember 18, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/904,991

```

```

; FILING DATE: June 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: CWR 2 149-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 861-5582
; TELEFAX: (216) 241-1666
; TELEX: (216) 980162
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 98
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: amino acids
; HYPOTHETICAL: irrelevant
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: mouse
; INDIVIDUAL ISOLATE: 8 day old mouse brain
; IMMEDIATE SOURCE:
; LIBRARY: lamda GT 10 and lamda GT11
; CLONE: synthesis of several clones
; PUBLICATION INFORMATION:
; AUTHORS: Moos, M.
; AUTHORS: Tacke, R.
; AUTHORS: Scherer, H.
; AUTHORS: Teplow, D.
; AUTHORS: Fruh, K.
; AUTHORS: Schachner, M.
; TITLE: Neural adhesion molecule L1 is a
; TITLE: member of the immunoglobulin
; TITLE: superfamily with binding domains
; TITLE: similar to fibronectin
; JOURNAL: NATURE
; VOLUME: 334
; ISSUE:
; PAGES: 701-703
; DATE: 1988
; US-08-341-843B-13

Query Match 11.7%; Score 65.5; DB 2; Length 98;
Best Local Similarity 27.9%; Pred. No. 0.22;
Matches 19; Conservative 12; Mismatches 26; Indels 11; Gaps 3;

QY 41 CQLEDEEVLSH-SPRLTNSAMRAGHLLDLKTRGKNGAIAFLE-----SLKFHNPDVYTLVT 95
Db 11 CQSDTSLLLHWQPLSHNGVLTGYLLSYHPVEGESKEQLFFNLSDPELRTHN-----LT 64

QY 96 GLOPDVDF 103
Db 65 NLNPDLOY 72

RESULT 7
US-08-427-497E-18
; Sequence 18, Application US/08427497E
; Patent No. 5969124
; GENERAL INFORMATION:
; APPLICANT: Lemmon, Vance
; TITLE OF INVENTION: A Method for Characterizing the
; TITLE OF INVENTION: Nucleotide Sequence of LICAM and
; Patent No. 5969124
; TITLE OF INVENTION: the Nucleotide Sequence
; TITLE OF INVENTION: Characterized Thereby
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; CITY: Cleveland

```


; LENGTH: 397 amino acids

DD 19Z MRGCRHSGVRITITPPRNP--

Db 192 MRGCRHSGVRIIPPRKAP--QPTRVTCRYLGKDKLAHPPLSEGEALASRILEMAPHGA 249

Wed Jan 22 10:56:35 2003

```
Qy 68 -----LKTRGKNGAIAFLES 82
Db 250 KFLGPVILEVPHFASLRGREREIVILRS 277

RESULT 12
US-08-847-429A-15
; Sequence 15, Application US/08847429A
; Patent No. 5827692
; GENERAL INFORMATION:
; APPLICANT: Tang, Liang
; APPLICANT: Blehm, E. Scot
; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/847,429A
; FILING DATE: 24-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: HW-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 409 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-847-429A-15

Query Match 11.7%; Score 65.5; DB 2; Length 409;
Best Local Similarity 25.0%; Pred. No. 1.9;
Matches 22; Conservative 12; Mismatches 33; Indels 21; Gaps 3;

Qy 14 MESHRRIVRCICPSRLTPYLQAKVLCQ-LDEEVVLSPLTNSAMRAGHLIDL----- 67
Db 192 MRGCRHSGVRIIIPPRKAP--QPTRVTCRYLGKDKLAHPPLPPLSEGEALASRILEMAPHGA 249

Qy 68 -----LKTRGKNGAIAFLES 82
Db 250 KFLGPVILEVPHFASLRGREREIVILRS 277

RESULT 13
US-09-065-474-15
; Sequence 15, Application US/09065474
; Patent No. 6063599
; GENERAL INFORMATION:
; APPLICANT: Tang, Liang
; APPLICANT: Blehm, E. Scot
; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 171
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,034
; FILING DATE: 21-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:

US-09-557-034-15
; Sequence 15, Application US/09557034
; Patent No. 6365569
; GENERAL INFORMATION:
; APPLICANT: Tang, Liang
; APPLICANT: Blehm, E. Scot
; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 171
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,034
; FILING DATE: 21-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:

US-09-065-474-15
; Sequence 15, Application US/09065474
; Patent No. 6063599
; GENERAL INFORMATION:
; APPLICANT: Tang, Liang
; APPLICANT: Blehm, E. Scot
; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 171
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/847,429A
; FILING DATE: 24-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: HW-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 409 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-847-429A-15

Query Match 11.7%; Score 65.5; DB 2; Length 409;
Best Local Similarity 25.0%; Pred. No. 1.9;
Matches 22; Conservative 12; Mismatches 33; Indels 21; Gaps 3;

Qy 14 MESHRRIVRCICPSRLTPYLQAKVLCQ-LDEEVVLSPLTNSAMRAGHLIDL----- 67
Db 192 MRGCRHSGVRIIIPPRKAP--QPTRVTCRYLGKDKLAHPPLPPLSEGEALASRILEMAPHGA 249

Qy 68 -----LKTRGKNGAIAFLES 82
Db 250 KFLGPVILEVPHFASLRGREREIVILRS 277

RESULT 14
US-09-557-034-15
; Sequence 15, Application US/09557034
; Patent No. 6365569
; GENERAL INFORMATION:
; APPLICANT: Tang, Liang
; APPLICANT: Blehm, E. Scot
; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 171
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,034
; FILING DATE: 21-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: 09/065,474
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: HW-5-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 409 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-557-034-15

Query Match 11.7%; Score 65.5; DB 4; Length 409;
Best Local Similarity 25.0%; Pred. No.1.9;
Matches 22; Conservative 12; Mismatches 33; Indels 21; Gaps 3;

QY 14 MESHRRIVRCICPSRLTPYLQAKVLCO-LDEEEVLSPLRTNSAMRAGHLDDL----- 67
Db 192 MRGCRHSGVRIIIPPRKAP--QPTRVTCRYLGKDKLAHPPLSEGEALASRIEMAPHGA 249

QY 68 -----LKTRGKNGAIAFLES 82
Db 250 KFLGPVILEVPHFASLRGREREIVILRS 277

RESULT 15
US-09-065-474-142
; Sequence 142, Application US/09065474
; Patent No. 6063599
; GENERAL INFORMATION:
; APPLICANT: Tang, Liang
; APPLICANT: Blehm, E. Scot
; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
; TITLE OF INVENTION: PROTEINS NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 171
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows; Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/065,474
; FILING DATE: 24-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: HW-5-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-09-065-474-142
```

```

Query Match 11.7%; Score 65.5; DB 3; Length 422;
Best Local Similarity 25.0%; Pred. No.1.9;
Matches 22; Conservative 12; Mismatches 33; Indels 21; Gaps 3;

QY 14 MESHRRIVRCICPSRLTPYLQAKVLCO-LDEEEVLSPLRTNSAMRAGHLDDL----- 67
Db 6 MRGCRHSGVRIIIPPRKAP--QPTRVTCRYLGKDKLAHPPLSEGEALASRIEMAPHGA 63

QY 68 -----LKTRGKNGAIAFLES 82
Db 64 KFLGPVILEVPHFASLRGREREIVILRS 91

Search completed: January 22, 2003, 08:57:27
Job time : 9.33468 secs
```

THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	560	100.0	139	9	US-10-032-159A-16	Sequence 16, Appl
2	560	100.0	1004	10	US-09-767-215-2	Sequence 2, Appl
3	560	100.0	1138	10	US-09-767-215-5	Sequence 5, Appl
4	484	86.4	92	9	US-10-032-159A-8	Sequence 18, Appl
5	270	48.2	1247	9	US-10-032-159A-8	Sequence 8, Appl
6	257	45.9	92	9	US-10-032-159A-2	Sequence 10, Appl
7	209	37.3	366	9	US-10-032-159A-2	Sequence 2, Appl
8	209	37.3	366	9	US-10-032-159A-2	Sequence 20, Appl
9	158	28.2	72	9	US-10-032-159A-4	Sequence 4, Appl
10	67.5	12.1	763	9	US-09-738-626-4454	Sequence 4454, Appl
11	67	12.0	431	12	US-10-078-929-200	Sequence 200, Appl
12	65.5	11.7	1260	9	US-10-024-918-27	Sequence 27, Appl
13	63	11.2	433	10	US-09-771-161A-229	Sequence 229, Appl
14	63	11.2	433	10	US-09-771-161A-230	Sequence 230, Appl
15	62.5	11.2	100	9	US-10-058-820-7	Sequence 7, Appl
16	62.5	11.2	461	9	US-09-712-363-285	Sequence 285, Appl
17	62.5	11.2	476	9	US-10-058-820-4	Sequence 4, Appl
18	62.5	11.2	553	9	US-10-058-820-3	Sequence 3, Appl
19	62.5	11.2	926	10	US-09-910-150-11	Sequence 11, Appl

```
; Sequence 2, Application US/09767215
; Patent No. US20020081636A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; FILE REFERENCE: 07334-142001
; CURRENT APPLICATION NUMBER: US/09/767,215
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/181,159
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1004
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-767-215-2

Query Match      100.0%; Score 560; DB 10; Length 1004;
Best Local Similarity 100.0%; Pred. No. 4.4e-56;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ALTALDEETLWEMESHRRHVRICPCPSRLTPYLROAKVLCQDDEEVLHSPRLTNSAMR 60
    |||||
Db 10 ALTALDEETLWEMESHRRHVRICPCPSRLTPYLROAKVLCQDDEEVLHSPRLTNSAMR 69
    |||||
Qy 61 AGHLLDLLKTRGKNGAIAFLSKFHNPDPVYTLVTGLQPDVDFSNFS 107
    |||||
Db 70 AGHLLDLLKTRGKNGAIAFLSKFHNPDPVYTLVTGLQPDVDFSNFS 116
    |||||

RESULT 3
US-09-767-215-5
; Sequence 5, Application US/09767215
; Patent No. US20020081636A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; FILE REFERENCE: 07334-142001
; CURRENT APPLICATION NUMBER: US/09/767,215
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/181,159
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1138
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-767-215-5

Query Match      100.0%; Score 560; DB 10; Length 1138;
Best Local Similarity 100.0%; Pred. No. 5.1e-56;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ALTALDEETLWEMESHRRHVRICPCPSRLTPYLROAKVLCQDDEEVLHSPRLTNSAMR 60
    |||||
Db 9 ALTALDEETLWEMESHRRHVRICPCPSRLTPYLROAKVLCQDDEEVLHSPRLTNSAMR 68
    |||||
Qy 61 AGHLLDLLKTRGKNGAIAFLSKFHNPDPVYTLVTGLQPDVDFSNFS 107
    |||||
Db 69 AGHLLDLLKTRGKNGAIAFLSKFHNPDPVYTLVTGLQPDVDFSNFS 115
    |||||

RESULT 4
US-10-032-159A-18
; Sequence 18, Application US/10032159A
; Patent No. US20020164703A1
; GENERAL INFORMATION:
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Reed, John C.
```

```
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,
; FILE REFERENCE: P-LJ 5100
; CURRENT APPLICATION NUMBER: US/10/032,159A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,457
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-159A-18

Query Match      86.4%; Score 484; DB 9; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.1e-48;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 BETLWEMESHRRHVRICPCPSRLTPYLROAKVLCQDDEEVLHSPRLTNSAMRAGHLLD 66
    |||||
Db 1 BETLWEMESHRRHVRICPCPSRLTPYLROAKVLCQDDEEVLHSPRLTNSAMRAGHLLD 60
    |||||
Qy 67 LLKTRGKNGAIAFLSKFHNPDPVYTLVTGLQ 98
    |||||
Db 61 LLKTRGKNGAIAFLSKFHNPDPVYTLVTGLQ 92
    |||||

RESULT 5
US-10-032-159A-8
; Sequence 8, Application US/10032159A
; Patent No. US20020164703A1
; GENERAL INFORMATION:
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,
; FILE REFERENCE: P-LJ 5100
; CURRENT APPLICATION NUMBER: US/10/032,159A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,457
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-159A-8

Query Match      48.2%; Score 270; DB 9; Length 1247;
Best Local Similarity 51.5%; Pred. No. 1e-22;
Matches 51; Conservative 22; Mismatches 26; Indels 0; Gaps 0;

Qy 6 DEETLWEMESHRRHVRICPCPSRLTPYLROAKVLCQDDEEVLHSPRLTNSAMRAGHLL 65
    |||||
Db 11 BEDALWENVECNHRLSRYPINPAKLTPLYRQCKVIDEODEVLNAPMLPSKINRAGRL 70
    |||||
Qy 66 DLLKTRGKNGAIAFLSKFHNPDPVYTLVTGLQPDVDF 104
    |||||
Db 71 DILHTKGQRYGVVFLSELYFYYPYLYKVTGKPTRRFS 109
    |||||

RESULT 6
US-10-032-159A-10
; Sequence 10, Application US/10032159A
; Patent No. US20020164703A1
; GENERAL INFORMATION:
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
```

RESULT 10
US-09-738-626-4454
Sequence 4454, Application US09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIICHO

; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4454
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4454

Query Match 12.1%; Score 67.5; DB 9; Length 763;
Best Local Similarity 30.4%; Pred. No. 9.1;
Matches 21; Conservative 13; Mismatches 28; Indels 7; Gaps 3;

QY 36 QAKVLCQDEEVHSPRLTNSAMRAGHLDDLKTRGKNGAIAF---LESCLKFHNPDVYT 92
|| : : : : ||| |||| | : : : :
DB 439 QAFIALHSENNQISFGQALLDAALGK---VDLLGARGKNAVIFKNEFLDALRSELPWTWN 495

QY 93 LVTGLQPOV 101
||||| : :
DB 496 EVTGL-PDI 503

RESULT 11
US-10-078-929-200
; Sequence 200, Application US/10078929
; Patent No. US20020152497A1
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni
; APPLICANT: Miao, Guo-Hua
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Sakai, Hajime
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Odell, Joan T.
; APPLICANT: Meyers, Blake
; APPLICANT: Thorpe, Catherine
; APPLICANT: Weng, Zude
; TITLE OF INVENTION: Nucleic Acid Fragments Encoding Proteins Involved in
; FILE REFERENCE: BBI357 US NA
; CURRENT APPLICATION NUMBER: US/10/078,929
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/566,394
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 60/133038
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: 60/133042
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: 60/133427
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133437
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133428
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133438
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133436
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/137667
; PRIOR FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 200
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Ipomoea batatas
US-10-078-929-200

Query Match 12.0%; Score 67; DB 12; Length 431;
Best Local Similarity 24.6%; Pred. No. 5;
Matches 32; Conservative 17; Mismatches 45; Indels 36; Gaps 6;

QY 7 EETLWMMESHRRIVRCICP-----SRITP-----YLRQAK-VLC 41
| : : : : | | | | | | | |
DB 248 EGSTYEILAAH---IWRACKARGLTDDQATKLYVATDGRSLRCLPPLPGYLCNVVFTAT 304

QY 42 QLDDEEVHSPRLTNSAMRAGHLDDLKTRGKNGAIAFLESCLKFHNPDVYTVTG---L 97
: : : : | | | | | : : : : | | : : : :
DB 305 PMAESGELQSEPLTNSAKRIHSALSALRMDDEYLRSLADFLEC-----QPDLSKLIRGSNYFA 360

QY 98 QPDVDFSNFS 107
| : : : : : : : : : :
DB 361 SPNLNINSWT 370

RESULT 12
US-10-024-918-27
; Sequence 27, Application US/10024918
; Patent No. US20020168718A1
; GENERAL INFORMATION:
; APPLICANT: Hubbell, Jeffrey
; APPLICANT: Schense, Jason
; APPLICANT: Zisch, Andreas
; APPLICANT: Hall, Heike
; TITLE OF INVENTION: ENZYME-MEDIATED MODIFICATION OF FIBRIN FOR TISSUE ENGINEERING
; FILE REFERENCE: CIT 2606 CIP
; CURRENT APPLICATION NUMBER: US/10/024,918
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 1260
; TYPE: PRT
; ORGANISM: Mus Musculus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (516)..(604)
; OTHER INFORMATION: the sixth Ig-like domain of the cell adhesion molecule L1
US-10-024-918-27

Query Match 11.7%; Score 65.5; DB 9; Length 1260;
Best Local Similarity 27.9%; Pred. No. 29;
Matches 19; Conservative 12; Mismatches 26; Indels 11; Gaps 3;

QY 41 COLDEEVH-H-SPRLTNSAMRAGHLDDLKTRGKNGAIAFLE-----SLKFHNPVTVTVT 95
|| : : : : | : : : : | : : : : | : : : :
DB 927 CQSDTSLLLHWQPLSHNGVLTGYLLSYHPVEGESKQLFFNLSDPELRTHN-----LT 980

QY 96 GLQPDVDF 103
| : : : :
DB 981 NLNPDLQY 988

RESULT 13
US-09-771-161A-229
; Sequence 229, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26

```

; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 229
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-229

Query Match      11.2%; Score 63; DB 10; Length 433;
Best Local Similarity 21.0%; Pred. No. 15;
Matches 30; Conservative 19; Mismatches 36; Indels 58; Gaps 7;

Qy 6 DEETLWEMESHRRIVRCI-----PSRLTPYLQAKVLCQL-DEEVLHSPRLT 55
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 EKQKMYWME-----YCVGCMQEMLDSPVKRFPVCQAHGYFCQLDGLGYLHSGIV 173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 56 NSAMRAGHLLDLLKTRG--KNGATAFLESL-----KFNPDV----- 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 HKDIKPGNL--LLTTGGTLKISDLGVAELHPFAADTCRTSQGSPAFQPPPEIANGLDTF 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 91 -----YTLVTGLQP 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 SGFKVDIWSAGVTLYNITGLYP 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-09-771-161A-230
; Sequence 230, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771.161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 230
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-230

Query Match      11.2%; Score 63; DB 10; Length 433;
Best Local Similarity 21.0%; Pred. No. 15;
Matches 30; Conservative 19; Mismatches 36; Indels 58; Gaps 7;

Qy 6 DEETLWEMESHRRIVRCI-----PSRLTPYLQAKVLCQL-DEEVLHSPRLT 55
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 EKQKMYWME-----YCVGCMQEMLDSPVKRFPVCQAHGYFCQLDGLGYLHSGIV 173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 56 NSAMRAGHLLDLLKTRG--KNGATAFLESL-----KFNPDV----- 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 HKDIKPGNL--LLTTGGTLKISDLGVAELHPFAADTCRTSQGSPAFQPPPEIANGLDTF 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 91 -----YTLVTGLQP 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 SGFKVDIWSAGVTLYNITGLYP 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
```

```

US-10-058-820-7
; Sequence 7, Application US/10058820
; Patent No. US20020155479A1
; GENERAL INFORMATION:
; APPLICANT: Bogan, Jonathan S.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: Expression Cloning Method
; FILE REFERENCE: 0399.2025-002
; CURRENT APPLICATION NUMBER: US/10/058,820
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/325,651
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 60/298,963
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/264,816
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-058-820-7

Query Match      11.2%; Score 62.5; DB 9; Length 100;
Best Local Similarity 29.3%; Pred. No. 2.6;
Matches 24; Conservative 12; Mismatches 29; Indels 17; Gaps 5;

Qy 8 ETLWEMESHRRIVRCI-CPSRLTPYLQAKVLCQDDEEVLHSPRLTNSAMRAGHLLD 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 35 QTLWELL-SHFPOIRECLQHPGGATP-----VCVYTRDEV-----TGEAALRGTTLQ 80
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 67 LLKTRGKNGAIAFLESKLFHNP 88
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 81 SLGLTGSATIRFV--MKCYDP 100
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Search completed: January 22, 2003, 08:52:45
Job time : 5.31452 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 8.19758 Seconds
(without alignments)
1254.807 Million cell updates/sec

Title: US-09-767-215-2_COPY_10_116
Perfect score: 560
Sequence: 1 ALTALDEFTLWEMESHRR.....PDVYTLVTGLQPDVDFSNFS 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73	13.0	210	2 S55671	hypothetical prote
2	73	13.0	265	2 H97280	hydroxyethylthiazo
3	72.5	12.9	1583	2 S59644	sister chromatid c
4	71	12.7	265	2 T35012	hypothetical prote
5	69.5	12.4	522	2 D72349	conserved hypotnet
6	67.5	12.1	270	2 T50952	transposase relate
7	67.5	12.1	315	2 T26673	hypothetical prote
8	67.5	12.1	460	2 E86922	probable FAD-link
9	67	12.0	1446	2 T13018	hypothetical prote
10	66.5	11.9	521	2 AB0269	anthranilate synth
11	66	11.8	853	2 D95365	NADH2 dehydrogenas
12	65.5	11.7	448	2 H70947	hypothetical prote
13	65.5	11.7	496	2 E83849	spore germination
14	65.5	11.7	1260	1 S05479	neural cell adhesi
15	64.5	11.5	320	2 E89264	pyruvate formate-1
16	64.5	11.5	532	2 H83493	hypothetical prote
17	64.5	11.5	554	2 S75969	hypothetical prote
18	64.5	11.5	618	2 T39654	hypothetical prote
19	64	11.4	221	2 A41333	glycerol metabolis
20	64	11.4	234	2 T33951	hypothetical prote
21	64	11.4	599	2 AB3375	beta-(1->2)glucan
22	64	11.4	2150	1 S27402	zinc finger protei
23	64	11.4	2150	2 T19450	hypothetical prote
24	63.5	11.3	235	2 H72623	hypothetical prote
25	63.5	11.3	537	1 YRHUB6	tyrosinase-related
26	63	11.2	263	2 A30227	hypothetical prote
27	63	11.2	297	2 AD2318	transcription regu
28	63	11.2	349	2 AF2130	iron(III) diglitrat
29	63	11.2	394	2 E71136	threonine synthase

30	63	11.2	728	2 S57142	hypothetical prote
31	63	11.2	867	2 H70411	alanine-tRNA ligas
32	63	11.2	1025	2 T44802	type I site-specif
33	63	11.2	1025	2 T09459	type I site-specif
34	62.5	11.2	195	2 G38824	hypothetical prote
35	62.5	11.2	356	2 S13221	GTP-binding regula
36	62.5	11.2	373	2 E82570	chorismate mutase
37	62.5	11.2	461	2 B70697	probable oxidoredu
38	62.5	11.2	513	2 S69181	protein disulfide-
39	62.5	11.2	1036	2 T31673	N-acetylglucosamin
40	62.5	11.2	1111	2 T01239	hypothetical prote
41	62.5	11.2	1168	2 T15890	hypothetical prote
42	62.5	11.2	1259	2 S36126	neural cell adhesi
43	62	11.1	259	2 G97127	flagellar motor co
44	62	11.1	271	1 A89124	probable phosphoe
45	62	11.1	298	2 H64402	molybdopterin bios

ALIGNMENTS

RESULT 1
S55671
hypothetical protein E10 - equine herpesvirus 2
C:Species: equine herpesvirus 2
C>Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 24-Nov-1999
C:Accession: S55671
R:Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.
J. Mol. Biol. 249, 520-528, 1995
A:Title: The DNA sequence of equine herpesvirus 2.
A:Reference number: S55594; MUID:95302501; PMID:7783207
A:Accession: S55671
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-210 <TEL>
A:Cross-references: GB:U00824; MID:g695172; PIDN:AAC13865.1; PID:g695250
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995
C:Superfamily: equine herpesvirus 2 hypothetical protein E10

Query Match 13.0%; Score 73; DB 2; Length 210;
Best Local Similarity 29.5%; Pred. No. 2.5;
Matches 28; Conservative 13; Mismatches 24; Indels 30; Gaps 5;
QY 5 LDEETLWEM-----MESHRRIVRCICPSRLTPYLKQAKVLQCLDDEEVL 49
Db 16 LTEEDINDYVERLCLEBELRVLLVSLKSKHK-----LDHLRAKKILSREDAEV- 63
QY 50 HSPRLTNSAMRAGHLLDLLKTRKNGAIAFLSLK 84
Db 64 -SSRAT-SRSRAGLLVDMQDHPHPR-GFQCLKESCK 95
RESULT 2
H97280
hydroxyethylthiazole Kinase ThiM/ThiK (FS1) CAC3095 [imported] - Clostridium acetobut
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: H97280
R:Nolling, J.; Brecon, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H97280
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-265 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK81035.1; PID:gl15026160; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3095
C:Superfamily: phosphomethylpyrimidine phosphate kinase

[illegible]

Db 180 ETL--LVLDHQHSTR-LQASLFTPDSEYQRLATRLQLSHQLQAPHPATSPVEMA 236
Qy 41 --COLDEEVHSPRLNSAMRAGHLLDLKTRGN-----GAIAFLESKFNPDVTVLV 94
Db 237 LQNSDEYCNVYSELQVAIREGEIQVPSRRFTLPCPSPLAAYQTLKDHNPSPYMF- 295
Qy 95 TGLQPDVDFSNF 106
Db 296 --FMQDNDFSFLF 305

RESULT 11
D95355
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain G NucG2 [imported] - Sinorhizobium m
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 03-Jun-2002
C:Accession: D95365
R:Barrett, M.J.; Jones, R.F.; Fisher, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilot
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: D95365
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>
A:CROSS-references: GB:AE006469; PIDN:AAK65486.1; PID:g14523958; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSYMA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: nuoG2
A:Genome: plasmid
C:Keywords: oxidoreductase

Query Match 11.8%; Score 66; DB 2; Length 853;
Best Local Similarity 30.8%; Pred. No. 76;
Matches 16; Conservative 12; Mismatches 24; Indels 0; Gaps 0;

Qy 29 RLTPYLROAKVLCQDEEVHSPRLNSAMRAGHLLDLKTRGNKGAIAFL 80
Db 470 RLVTSLKEGRSVTLVSVDLRLSPARKTLEQLGNLQLLLRLGKPSLQFL 521

RESULT 12
H70947
hypothetical protein Rv3170 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: H70947
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70947
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-448 <COL>
A:CROSS-references: GB:AL021646; GB:AL123456; NID:g3242278; PIDN:CAA16635.1; PID:e124876
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3170

Query Match 11.7%; Score 65.5; DB 2; Length 448;
Best Local Similarity 31.5%; Pred. No. 40;
Matches 23; Conservative 12; Mismatches 27; Indels 11; Gaps 3;

Qy 5 LDEETLWEMSHRRIVRICPSR-LTPYLROAKVLCQDEEVHSPRLNSAMRAGH 63
Db 144 LDDVSLGELW-----RLVTRSSSRNLMAIMTRVTCGCPDDVSMILHAARYVRAAGGLDR 198

Qy 64 LDLLKTRGNKA 76
Db 199 LLDV-----KNGA 206

RESULT 13
E83849
spore germination protein BH1597 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: E83849
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: E83849
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-496 <STO>
A:CROSS-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA05316.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1597
C:Superfamily: spore germination protein gerBA

Query Match 11.7%; Score 65.5; DB 2; Length 496;
Best Local Similarity 27.1%; Pred. No. 45;
Matches 23; Conservative 20; Mismatches 21; Indels 21; Gaps 5;

Qy 30 LTPYL-ROAKVLCQDEE-----EVLHSPRLNSAMRAGHLLDLKTRGNKGAIAFLES 82
Db 61 VTPYLKKEAKTLEELIEDVPIONKEISDDPQOIKTVMQGFV--LLRKKGEGHEVALIE- 117

Qy 83 LKFHNPVDVYTLVTGLQ---PDVDFS 104
Db 118 -----VTMDSGREVQRPEVFS 134

RESULT 14
S05479
neural cell adhesion molecule L1 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S05479; B60850; S22167
R:Moos, M.; Tacke, R.; Scherer, H.; Teplov, D.; Frueh, K.; Schachner, M.
Nature 334, 701-703, 1988
A:Title: Neural adhesion molecule L1 as a member of the immunoglobulin superfamily wi
A:Reference number: S05479; MUID:88318924; PMID:3412448
A:Accession: S05479
A:Molecule type: mRNA
A:Residues: 1-1260 <MOO>
A:CROSS-references: EMBL:X12875; NID:g53336; PIDN:CAA31368.1; PID:g53337
A:Note: the authors translated the codon CCT for residue 166 as Leu, ACT for residue
A:Note: part of this sequence, including the amino end of the mature protein, was con
R:Rathjen, F.G.; Wolff, J.M.; Frank, R.; Bonhoeffer, F.; Rutishauser, U.
J. Cell Biol. 104, 343-353, 1987
A:Title: Membrane glycoproteins involved in neurite fasciculation.
A:Reference number: A60850; MUID:87109457; PMID:3805123
A:Accession: B60850
A:Molecule type: protein
A:Residues: 20-28, 'XX', 31-36 <RAT>
R:Kohl, A.; Giese, K.P.; Mohajeri, M.H.; Montag, D.; Moos, M.; Schachner, M.
submitted to the EMBL Data Library, December 1991
A:Description: Analysis of promoter activity and 5' genomic structure of the neural c
A:Reference number: S22167

Search completed: January 22, 2003, 08:54:27
Job time : 11.1976 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 4.1707 Seconds
(without alignments)
1064.082 Million cell updates/sec

Title: US-09-767-215-2_COPY_10_116

Perfect score: 560

Sequence: 1 ALTADETLWEMESHRR.....PDVYTLVTGLQPDVDFSNFS 107

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	560	100.0	1004	1 CARE_HUMAN	Q9bx16 homo sapien
2	470	83.9	999	1 CARE_MOUSE	Q99kf0 mus musculus
3	270	48.2	1147	1 CARB_HUMAN	Q9bx17 homo sapien
4	252	45.0	1021	1 CARA_MOUSE	P58660 mus musculus
5	246	43.9	1032	1 CARA_HUMAN	Q9bwt7 homo sapien
6	217	38.8	536	1 CAR9_RAT	Q9epv0 rattus norv
7	209	37.3	536	1 CAR9_HUMAN	Q9h257 homo sapien
8	81	14.5	233	1 BCLA_HUMAN	Q95999 h b cell ly
9	81	14.5	233	1 BCLA_RAT	Q9qyn5 rattus norv
10	78	13.9	233	1 BCLA_MOUSE	Q9z0h7 m b cell ly
11	72.5	12.9	1583	1 MISA_SCHPO	Q09725 schizosach
12	66	11.8	853	1 NUG2_RHIME	P56914 rhizobium m
13	65.5	11.7	454	1 AOPH_MYCTU	Q53320 mycobacteri
14	65.5	11.7	685	1 STM1_HUMAN	Q13586 homo sapien
15	65.5	11.7	1260	1 CAML_MOUSE	P11627 mus musculus
16	65	11.6	352	1 TPO_CANFA	P42705 canis famil
17	65	11.6	578	1 V022_FOWPV	Q915h9 fowlpox vir
18	65	11.6	725	1 GCP5_MACFA	Q95k09 macaca fasc
19	65	11.6	1024	1 GCP5_HUMAN	Q96rt8 homo sapien
20	64.5	11.5	554	1 Y514_SYNY3	Q53470 synecocyst
21	64	11.4	221	1 AGMR_PSEAE	P29369 pseudomonas
22	64	11.4	260	1 RIFF_ARYMD	O52547 amycolatops
23	64	11.4	2150	1 SOC3_CAEEL	P34706 caenorhabdi
24	63.5	11.3	537	1 TYR1_HUMAN	P17643 homo sapien
25	63.5	11.3	681	1 CAQ2_MOUSE	Q9qxd1 mus musculus
26	63	11.2	433	1 ST11_HUMAN	Q15831 homo sapien
27	63	11.2	728	1 YJ89_YEAST	P47156 saccharomyc
28	63	11.2	867	1 SYA_AQUAE	O67323 aquifex aeo
29	62.5	11.2	356	1 GBA2_CAEEL	P22454 caenorhabdi
30	62.5	11.2	513	1 PDI_MAIZE	P52588 zea mays (m
31	62.5	11.2	1036	1 OGTL_HUMAN	O15294 homo sapien
32	62.5	11.2	1036	1 OGTL_RAT	P56558 rattus norv
33	62.5	11.2	1259	1 CAML_RAT	Q05695 rattus norv

RESULT 1

CARE_HUMAN

ID	CARE_HUMAN	STANDARD;	PRT;	1004 AA.
AC	Q9BX16; Q9BVB5;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Caspase recruitment domain protein 14 (CARD-containing MAGUK protein 2) (Carma 2).			
GN	CARD14 OR CARMA2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21192234; PubMed=11278692;			
RA	Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,			
RA	Srinivasula S.M., Merriam S., DiStefano P.S., Alnemri E.S.;			
RT	"CARD11 and CARD14 are novel caspase recruitment domain			
RT	(CARD)/membrane-associated guanylate kinase (MAGUK) family members			
RT	that interact with Bcl10 and activate NF-kappaB.";			
RL	J. Biol. Chem. 276:11877-11882(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21255663; PubMed=11356195;			
RA	Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;			
RT	"Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10			
RT	phosphorylation and NF-kappaB activation.";			
RL	FEBS Lett. 496:121-127(2001).			
RN	[3]			
RP	ERRATUM.			
RA	Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;			
RL	FEBS Lett. 505:198-198(2001).			
RN	[4]			
RP	SEQUENCE OF 1-740 FROM N.A.			
RC	TISSUE=Cervix, and Colon;			
RA	Strausberg R.;			
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the			
CC	phosphorylation of Bcl10.			
CC	-!- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD			
CC	interaction.			
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic.			
CC	-!- TISSUE SPECIFICITY: Expressed in placenta. Also detected in HeLa			
CC	S3 cells, but not in the other cancer cell lines tested.			
CC	-!- SIMILARITY: CONTAINS 1 CARD DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.			
CC	-!- CAUTION: Supposed to contain a SH3 domain which is not detected by			
CC	PROSITE, Pfam or SMART.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			

ALIGNMENTS

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF322642; AAG53403.1; -;
DR EMBL; AY032927; AAK54453.1; -;
DR EMBL; BC018142; AAH18142.1; -;
DR EMBL; BC001326; AAH01326.1; ALT_INIT.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR SMART; SM00072; GuK; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0209; CARD; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Coiled coil.
FT DOMAIN 15 107 CARD.
FT DOMAIN 128 409 COILED COIL (POTENTIAL).
FT DOMAIN 568 658 PDZ.
FT DOMAIN 858 990 GUANYLATE KINASE.
FT CONFLICT 619 671 DYASEPLFKAVLEDTILEEAVGLLRVVDGFCCLSVKNTD
FT GYKRLQLDEAK -> SRARELLSPGLLMGTVAAGGVQAD
FT FTSPRCRSTLGSALSADYKRSALH (IN REF. 2;
FT AAH01326).
SQ SEQUENCE 1004 AA; 113299 MW; 7469B8B56BE06073 CRC64;

Query Match 100.08; Score 560; DB 1; Length 1004;
Best Local Similarity 100.08; Pred. No. 3.5e-52;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALTADEETLWEMESHRRIVRCICPSRLTPYLQAKVLCOLDEEEVLHSPRLTNSAMR 60
DB 10 ALTADEETLWEMESHRRIVRCICPSRLTPYLQAKVLCOLDEEEVLHSPRLTNSAMR 69
QY 61 AGHLLDLKTRGKNGAIAFLESLEKFNPDVYTLVTGLQPDVDFSNFS 107
DB 70 AGHLLDLKTRGKNGAIAFLESLEKFNPDVYTLVTGLQPDVDFSNFS 116

RESULT 2
CARE_MOUSE STANDARD; PRT; 999 AA.
AC Q99KF0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 14 (Bcl10-interacting MAGUK protein
DE 2) (Bimp2).
GN CARD14 OR BIMP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21391892; PubMed=11387339;
RA McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A.,
RA Li Q., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W.,
RA Nunez G.;
RT "Bimp1, a MAGUK family member linking protein kinase C activation to
RT Bcl10-mediated NF-kappa B induction.";
RT J. Biol. Chem. 276:30589-30597 (2001).
RN [2]
RP SEQUENCE OF 82-743 FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC -1- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD
CC interaction (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: CONTAINS 1 CARD DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -1- CAUTION: Supposed to contain a SH3 domain which is not detected by
CC PROSITE, Pfam or SMART.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF363457; AAK60137.1; -;
DR EMBL; BC004692; AAH04692.1; -;
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR PROSITE; PS0209; CARD; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Coiled coil.
FT DOMAIN 15 107 CARD.
FT DOMAIN 125 411 COILED COIL (POTENTIAL).
FT DOMAIN 572 655 PDZ.
FT DOMAIN 854 986 GUANYLATE KINASE.
FT CONFLICT 736 743 QAQQQLLA -> HLEEDHRS (IN REF. 2).
SQ SEQUENCE 999 AA; 113496 MW; D18350DA12430255 CRC64;

Query Match 83.9%; Score 470; DB 1; Length 999;
Best Local Similarity 85.8%; Pred. No. 1.6e-42;
Matches 91; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 2 LTALDEETLWEMESHRRIVRCICPSRLTPYLQAKVLCOLDEEEVLHSPRLTNSAMRA 61
DB 11 LTALDEEMLDMLSHRCIRVQSCPSRLTPYLQAKVLCOLDEEEVLHSPRLTNSAMRV 70
QY 62 GHLLDLKTRGKNGAIAFLESLEKFNPDVYTLVTGLQPDVDFSNFS 107
DB 71 GHLLDLKTRGKNGAIAFLESLEKFNPDVYTLVTGLQSDIDSTFS 116

RESULT 3
CARE_HUMAN STANDARD; PRT; 1147 AA.
AC Q9BXI7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 11 (CARD-containing MAGUK protein
DE 3) (Carma 1).
GN CARD11 OR CARMA1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21192234; PubMed=11278692;
RA Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA Srinivasula S.M., Merriam S., Distefano P.S., Alnemri E.S.;
RT "CARD11 and CARD14 are novel caspase recruitment domain
RT (CARD)/membrane-associated guanylate kinase (MAGUK) family members
RT that interact with Bcl10 and activate NF-kappaB.";
RL J. Biol. Chem. 276:11877-11882(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21255663; PubMed=11356195;
RA Gaide O., Martinon F., Mischeau O., Bonnet D., Thome M., Tschopp J.;
RT "Carma1, a CARD-containing binding partner of Bcl10, induces Bcl10
RT phosphorylation and NF-kappaB activation.";
RL FEBS Lett. 496:121-127(2001).

```

RN  [3]
RA  ERRATUM.
RA  Gaide O., Martinon F., Mischeau O., Bonnet D., Thome M., Tschopp J.;
RL  FEBS Lett. 505:198-198(2001).
CC  -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the
CC  phosphorylation of Bcl10.
CC  -!- SUBUNIT: CARD11 and Bcl10 bind to each other by CARD-CARD
CC  interaction.
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC  -!- TISSUE SPECIFICITY: Detected in adult peripheral blood leukocytes,
CC  thymus, spleen and liver. Also found in promyelocytic leukemia HL-
CC  60 cells, chronic myelogenous leukemia K562 cells, Burkitt's
CC  lymphoma Raji cells and colorectal adenocarcinoma SW480 cells. Not
CC  detected in HeLa S3, Molt-4, A549 and G431 cells.
CC  -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC  -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC  -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC  -!- CAUTION: Supposed to contain a SH3 domain which is not detected by
CC  PROSITE, Pfam or SMART.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: AF322641; AAG53402.1; -.
DR  Genew: HGNC:16393; CARD11.
DR  InterPro: IPR000619; Guanylate_kin.
DR  SMART: IPR001478; PDZ.
DR  SMART: SM00228; PDZ; 1.
DR  PROSITE: PS50209; CARD; 1.
DR  PROSITE: PS00856; GUANYLATE_KINASE.1; FALSE_NEG.
DR  PROSITE: PS50052; GUANYLATE_KINASE_2; FALSE_NEG.
DR  PROSITE: PS50106; PDZ; FALSE_NEG.
KW  Coiled coil.
FT  DOMAIN 11 103 CARD.
FT  DOMAIN 123 442 COILED COIL (POTENTIAL).
FT  DOMAIN 673 748 PDZ.
FT  DOMAIN 966 1133 GUANYLATE KINASE.
FT  CONFLICT 808 808 P->L (IN REF. 2).
SQ  SEQUENCE 1147 AA; 132641 MW; 913A4B015D2B36CC CRC64;

Query Match 48.2%; Score 270; DB 1; Length 1147;
Best Local Similarity 51.5%; Pred. No. 5,1e-21;
Matches 51; Conservative 22; Mismatches 26; Indels 0; Gaps 0;

Qy 6 DEETLWEMESHRRIVRCICPSRLTPYLROAKVLCQDDEEVLSRLTNSAMRAGHLL 65
Db 11 EEDALWENVECNHLSRYINPAKLTPLYLOCKVIDEODEVLNAPMLPSKINRAGLL 70

Qy 66 DLLKTRGNGAIAFLESLEKFNHPDVTYLTGLOPDVFS 104
Db 71 DILHTKGORGIVVFLESLEFYYPELYLVLTGKEPTRRS 109

RESULT 4
CARA_MOUSE
ID CARA_MOUSE STANDARD; PRT; 1021 AA.
AC P38660;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Caspase recruitment domain protein 10 (Bcl10-interacting MAGUK protein
DE 1) (Bimpr).
DE CARD10 OR BIMPL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=21391892; PubMed=11387339;
RA McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A.,
RA Li O., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W.,
RA Nunez G.;
RA "Bimpr", a MAGUK family member linking protein kinase c activation to
RT Bcl10-mediated NF-kappa B induction."
RL J. Biol. Chem. 276:30589-30597(2001).
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC -!- SUBUNIT: CARD10 and Bcl10 bind to each other by CARD-CARD
CC interaction. They both participate in a complex with MALT1, where
CC MALT1 binds to Bcl10.
CC -!- TISSUE SPECIFICITY: Highly expressed in kidney, heart followed by
CC brain, lung, liver, skeletal muscle and testis.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- CAUTION: Supposed to contain a SH3, a PDZ and a guanylate kinase-
CC like domain. But none of these 3 domains are detected by PROSITE,
CC Pfam or SMART.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF363456; AAK60136.1; -.
DR PROSITE: PS50209; CARD; 1.
KW Coiled coil.
FT DOMAIN 23 115 CARD.
FT DOMAIN 138 450 COILED COIL (POTENTIAL).
FT DOMAIN 558 565 POLY-SER.
SQ SEQUENCE 1021 AA; 114413 MW; 4811A09BDB8F792C CRC64;

Query Match 45.0%; Score 252; DB 1; Length 1021;
Best Local Similarity 51.1%; Pred. No. 3,8e-19;
Matches 48; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

Qy 6 DEETLWEMESHRRIVRCICPSRLTPYLROAKVLCQDDEEVLSRLTNSAMRAGHLL 65
Db 23 EEDALWENVECNHLSRYINPAKLTPLYLOCKVIDEODEEVLSYRPPCRANRTGRLI 82

Qy 66 DLLKTRGNGAIAFLESLEKFNHPDVTYLTGLOP 99
Db 83 DILRCRKGKRGPEAFLEAFLEFYYPPEHFTLLTGQEP 116

RESULT 5
CARA_HUMAN
ID CARA_HUMAN STANDARD; PRT; 1032 AA.
AC Q9BW77; Q9UGR5; Q9UGR6; Q9Y3H0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Caspase recruitment domain protein 10 (CARD-containing MAGUK protein
DE 3) (Carma 3).
DE CARD10 OR CARMA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21292987; PubMed=11259443;
RA Wang L., Guo Y., Huang W.-J., Ke X., Poyet J.-L., Manji G.A.,
RA Merriam S., Glucksman M.A., Distefano P.S., Alnemri E.S., Bertin J.;
RA "CARD10 is a novel caspase recruitment domain/membrane-associated
RT guanylate kinase family member that interacts with Bcl10 and activates
RT NF-kappa B."
RL J. Biol. Chem. 276:21405-21409(2001).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=21255663; PubMed=11356195;
RA Galde O., Martinon F., Micheau O., Bonnet D., Thome M., Tschoep J.;
RT "Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10
RT phosphorylation and NF-kappaB activation.";
RL FEBS Lett. 496:121-127(2001).
RN [3]
RP ERRATUM.
RA Galde O., Martinon F., Micheau O., Bonnet D., Thome M., Tschoep J.;
RL FEBS Lett. 505:198-198(2001).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burdill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,
RA McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelman L.,
RA Kim U.-J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC -!- SUBUNIT: CARD10 and Bcl10 bind to each other by CARD-CARD
CC interaction. They both participate in a complex with MALTL1, where
CC MALTL1 binds to Bcl10 (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- TISSUE SPECIFICITY: Detected in adult heart, kidney and liver;
CC in fetal lung, liver and kidney.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- CAUTION: Supposed to contain a SH3, a PDZ and a guanylate kinase-
CC like domain. But none of these 3 domains are detected by PROSITE,
CC Pfam or SMART.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to various
CC gene identification problems.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AF311288; AAG28791.1; -.
CC InterPro: IPR001315; CARD.
CC PROSITE: PS50209; CARD; 1.
-----
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AY028896; AAK26165.1; -.
CC EMBL: AY032928; AAK54454.1; -.
CC EMBL: AL049851; CAB63075.1; ALT_SEQ.
CC EMBL: AL049851; CAB63076.1; ALT_SEQ.
CC EMBL: AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE: PS50209; CARD; 1.
KW Coiled coil.
FT DOMAIN 23 115 CARD.
FT DOMAIN 138 456 COILED COIL (POTENTIAL).
FT DOMAIN 567 574 POLY-SER.
FT CONFLICT 289 289 Q -> R (IN REF. 4).
FT CONFLICT 917 917 K -> Q (IN REF. 4; CAB63075).
FT CONFLICT 932 932 R -> L (IN REF. 4).
SQ SEQUENCE 1032 AA; 115946 MW; 8377319AB82A0949 CRC64;
Query Match 43.9%; Score 246; DB 1; Length 1032;
Best Local Similarity 48.9%; Pred. No. 1.7e-18;
Matches 46; Conservative 21; Mismatches 27; Indels 0; Gaps 0;
QY 6 DEETLWMESHRRHRIYRCIPSLPYLRQAKVLCQDEEVLHSPRLTNSAMRAGHLL 65
Db 23 EEDALWERIEGVHRLARALNPAKLTPYLRQCRVIDQDEEVLSTYRFPQVRNRTGLM 82
QY 66 DLLKTRGKNGAIFLESLEKPNPDVYTLVGLQP 99
Db 83 DILRCRGRGYEAFLEALEFYYPHEFTLLTGQEP 116
RESULT 6
CAR9_RAT STANDARD; PRT; 536 AA.
AC Q9EPY0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 9 (rCARD9).
GN CARD9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=20576269; PubMed=11053425;
RA Bertin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D.,
RA Poyet J.-L., Merriam S., Du M.-Q., Dyer M.J.S., Robison K.E.,
RA Distefano P.S., Alnemri E.S.;
RA "CARD9 is a novel caspase recruitment domain-containing protein that
RA interacts with Bcl10/CLAP and activates NF-kappaB via Bcl10 (By similarity)".
RL J. Biol. Chem. 275:41082-41086(2000).
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and Bcl10 bind to each other by
CC -!- SUBUNIT: Self-associates. CARD9 and Bcl10 bind to each other by
CC CARD-CARD interaction (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AF311288; AAG28791.1; -.
CC InterPro: IPR001315; CARD.
CC PROSITE: PS50209; CARD; 1.
-----
```

```
KW Coiled coil. 6 98 CARD.
FT DOMAIN 117 277 COILED COIL (POTENTIAL).
FT DOMAIN 303 420 COILED COIL (POTENTIAL).
SQ SEQUENCE 536 AA; 62631 MW; 6F33089CB7E6BAC9 CRC64;

Query Match 38.8%; Score 217; DB 1; Length 536;
Best Local Similarity 45.5%; Pred. No. 1e-15;
Matches 45; Conservative 17; Mismatches 37; Indels 0; Gaps 0;

QY 6 DEETLWEMSHRHVRICPSRLTPYLROAKVLCQDDEEVLSHSPRLTNSAMRAGHLL 65
DQ 6 NDDECSALESRVLIISVIDPSRITPYLRQCKVLNPDDEEQVLSDPNLVIRKRVGVLL 65
QY 66 DLLKTRGKNGAIAFLSKLHNPDPVYTLVTGLQPDVDFS 104
DQ 66 DILQRTGHKGYYVAFLESLEYYPQLYKVTGKEPARVES 104

RESULT 7
CAR9_HUMAN
ID CAR9_HUMAN STANDARD; PRT; 536 AA.
AC Q9H257; Q9H854;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 9 (hCARD9).
GN CARD9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20576268; PubMed=11053425;
RA Bertin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D.,
RA Poyet J.-L., Merriam S., Du M.-Q., Dyer M.J.S., Robison K.E.,
RA DiStefano P.S., Alnemri E.S.;
"CARD9 is a novel caspase recruitment domain-containing protein that
interacts with Bcl10/CLAP and activates NF-kappa B.";
J. Biol. Chem. 275:41082-41086(2000).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Retinoblastoma;
RA Ishigai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Ninomiya K., Iwayanagi T.;
"NEDO human cDNA sequencing project.";
Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE OF 1-492 FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Activates NF-kappaB via Bcl10.
CC -!- SUBUNIT: Self-associates. CARD9 and Bcl10 bind to each other by
CC CARD-CARD interaction.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Highly expressed in spleen. Also detected in
CC liver, placenta, lung, peripheral blood leukocytes and in brain.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 360.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF311287; AAG28790.1; -
DR EMBL; AK024001; BAB14766.1; ALT_FRAME.
DR EMBL; BC008877; AAH08877.1; -
DR Genew; HGNC:16391; CARD9.
DR InterPro; IPR001315; CARD.
DR PROSITE; PS50209; CARD; 1.
KW Coiled coil.
FT DOMAIN 6 98 CARD.
FT DOMAIN 117 277 COILED COIL (POTENTIAL).
FT DOMAIN 332 419 COILED COIL (POTENTIAL).
FT CONFLICT 12 12 N -> S (IN REF. 3).
FT CONFLICT 482 492 LSSGPPPEKER -> PAGLPGIGAYC (IN REF. 3).
SQ SEQUENCE 536 AA; 62267 MW; 6EB1835315B83DE5 CRC64;

Query Match 37.3%; Score 209; DB 1; Length 536;
Best Local Similarity 44.4%; Pred. No. 7.5e-15;
Matches 44; Conservative 16; Mismatches 39; Indels 0; Gaps 0;

QY 6 DEETLWEMSHRHVRICPSRLTPYLROAKVLCQDDEEVLSHSPRLTNSAMRAGHLL 65
DQ 6 NDDECNVLEGEFRVLTSTVIDPSRITPYLRQCKVLNPDDEEQVLSDPNLVIRKRVGVLL 65
QY 66 DLLKTRGKNGAIAFLSKLHNPDPVYTLVTGLQPDVDFS 104
DQ 66 DILQRTGHKGYYVAFLESLEYYPQLYKVTGKEPARVES 104

RESULT 8
BCLA_HUMAN
ID BCLA_HUMAN STANDARD; PRT; 233 AA.
AC O95999;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE B cell lymphoma/leukemia 10 (B-cell CLL/lymphoma 10) (Bcl-10) (CED-
DE 3/ICH-1 prodomain homologous E10-like regulator) (CIPER) (CARD-
DE containing molecule enhancing NFkappaB) (Cellular homolog of vCARMEN)
DE (cCARMEN) (Mammalian CARD-containing adapter molecule E10) (mE10)
DE (Cellular-E10) (c-E10) (CARD-like apoptotic protein) (hCLAP).
GN BCL10 OR CIPER OR CLAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.; VARIANT FOLLICULAR LYMPHOMA GLU-210 DEL, VARIANT
RP MESOTHELIOMA ILE-52, AND VARIANTS GERM CELL TUMOR GLY-58 AND PHE-218.
RC TISSUE=Lymphoma;
RX MEDLINE=99142601; PubMed=9989495;
RA Willis T.G., Jadavel D.M., Du M.-O., Peng H., Perry A.R.,
RA Abdul-Rauf M., Price H., Karan L., Majekodunmi O., Wlodarska I.,
RA Pan L., Crook T., Hamoudi R., Isaacson P., Dyer M.J.S.;
"Cell10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and
mutated in multiple tumor types.";
Cell 96:35-45(1999).
[2]
RP SEQUENCE FROM N.A.; AND MUTAGENESIS OF LEU-41 AND GLY-78.
RX MEDLINE=99214545; PubMed=10187770;
RA Koseki T., Inohara N., Chen S., Carrio R., Merino J., Hottiger M.O.,
RA Nabel G.J., Nunez G.;
"CIPER, a novel NF kappaB-activating protein containing a caspase
recruitment domain with homology to Herpesvirus-2 protein E10.";
J. Biol. Chem. 274:9955-9961(1999).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99214546; PubMed=10187771;
RA Thome M., Martinon F., Hofmann K., Rubio V., Steiner V., Schneider P.,
RA Matmann C., Tschopp J.;
"Equine herpesvirus-2 E10 gene product, but not its cellular
homologue, activates NF-kappaB transcription factor and c-Jun
```


KW	Complete proteome.
FT	METAL 34 34 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT	METAL 45 45 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT	METAL 48 48 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT	METAL 98 98 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT	METAL 101 101 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT	METAL 107 107 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT	METAL 148 148 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT	METAL 151 151 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT	METAL 154 154 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT	METAL 198 198 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT	CONFLICT 211 211 K -> R (IN REF. 1).
FT	CONFLICT 285 285 T -> A (IN REF. 1).
FT	CONFLICT 452 452 F -> L (IN REF. 1).
FT	CONFLICT 459 459 T -> I (IN REF. 1).
FT	CONFLICT 544 544 S -> G (IN REF. 1).
FT	CONFLICT 554 554 R -> K (IN REF. 1).
FT	CONFLICT 679 679 R -> Q (IN REF. 1).
FT	CONFLICT 734 734 A -> G (IN REF. 1).
FT	CONFLICT 756 756 H -> R (IN REF. 1).
SQ	SEQUENCE 853 AA; 93345 MW; CDF867CE3BD11A0B CRC64;
 Query Match 11.8%; Score 66; DB 1; Length 853;	
Best Local Similarity 30.8%; Pred.No. 27;	
Matches 16; Conservative 12; Mismatches 24; Indels 0; Gaps 0;	
QY	29 RLAPYLRAQKVLQOLDEEVLHSPRLNSAMRAGHLDDLKTRKGNAIAFL 80 : :: : : :: : : :: : ::
Db	470 RLVSTLKEGRSVLLVSVDLLRSPLARKTLEQLGNLLQLLRLLGKESLQFL 521 : :: : : :: : : :: : ::
RESULT 13	
AOFH_MYCTU	STANDARD; PRT; 454 AA.
ID AC	C53320;
DT	15-DEC-1998 (Rel. 37, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Putative flavin-containing monooxime oxidase Rv3170 (EC 1.4.3.*).
GN	Rv3170 OR MF3259 OR MV014.14.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OX	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;	
RN	[1]
RC	SEQUENCE FROM N.A.
RP	STRAIN=R37RV;
RX	MEDLINE=98295987; PubMed=9634230;
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.; "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";
RL	Nature 393:537-544 (1998).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CDC 1551 / Oshkosh;
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.; "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC	- COFACTOR: FAD (POTENTIAL).
CC	- - SIMILARITY: BELONGS TO THE FLAVIN MONOOXIME OXIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AL021646; CAAL16635.1; ALT_INIT.
DR EMBL: AE007139; AAK47598.1; -.
DR HSSP: P27338; IGOS.
DR TIGR: MT3259; -.
DR TubercuList; RV3170; -.
DR InterPro; IPR000759; Adrnx_reductase.
DR InterPro; IPR001613; Amineoxid.fl.
DR InterPro; IPR002937; Amino_oxidase.
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF01593; Amino_oxidase; 1.
DR PRINTS; PR00419; ADXRDTASE.
DR PRINTS; PR00757; AMINEOXDASEF.
KW Hypothetical protein; Oxidoreductase; Flavoprotein; FAD;
KW Complete proteome.
FT NP_BIND 14 69 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 454 AA; 49136 MW; 6C1AEB97FB2F435F CRC64;

Query Match 11.7%; Score 65.5; DB 1; Length 454;
Best Local Similarity 31.5%; Pred. No. 15;
Matches 23; Conservative 12; Mismatches 27; Indels 11; Gaps 3;

QY 5 LDETLWEMESHRRIVRCIPSR-LPFLYLRQAKVLCQDEEVLHSPRLTNSAMRAGH 63
DB 150 LDVSLGSEWL-----RLVRATSSRNLMATRVTWGCEPDDVSLHAARYVRAAGGLDR 204
QY 64 LLDLLKTRCKNGA 76
DB 205 LDV-----KNGA 212

RESULT 14
ID STIML_HUMAN STANDARD; PRT; 685 AA.
AC Q13586;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Stromal interaction molecule 1 precursor.
GN STIM1 OR GOK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta, and Fetal liver;
RX MEDLINE=97079692; PubMed=8921403;
RA Parker N.J., Begley C.G., Smith P.J., Fox R.M.;
RT "Molecular cloning of a novel human gene (D11S4896E) at chromosomal
RT region 11p15.5";
RL Genomics 37:253-256(1996).
RN [2]
RP GLYCOSYLATION, PHOSPHORYLATION, SUBCELLULAR LOCATION, AND TISSUE
RP SPECIFICITY.
RX MEDLINE=20461006; PubMed=11004585;
RA Manji S.S., Parker N.J., Williams R.T., van Stekelenburg L.,
RA Pearson R.B., Dziadek M., Smith P.J.;
RT STIM1: a novel phosphoprotein located at the cell surface";
RL Biochim. Biophys. Acta 1481:147-155(2000).
RN [3]
RP TISSUE SPECIFICITY, AND SUBUNIT.
RX MEDLINE=21356314; PubMed=11463338;
RA Williams R.T., Manji S.S.M., Parker N.J., Hancock M.S.,
RA Van Stekelenburg L., Eid J.-P., Senior P.V., Kazenwadel J.S.,
RA Shandala T., Saint R., Smith P.J., Dziadek M.A.;

RT "Identification and characterization of the STIM (stromal interaction
RT molecule) gene family: coding for a novel class of transmembrane
RT proteins";
RL Biochem. J. 357:673-685(2001).
RN [4]
RP DISEASE.

RX MEDLINE=98021968; PubMed=9377559;
RA Sabbioni S., Barbanti-Brodano G., Croce C.M., Negrini M.;
RT "GOK: a gene at 11p15 involved in rhabdomyosarcoma and rhabdoid tumor
RT development";
RL Cancer Res. 57:4493-4497(1997).
CC FUNCTION: Possible adhesion molecule with a role in early
CC hematopoiesis by mediating attachment to stromal cells. Influences
CC the survival and/or proliferation of B cell precursors. Binding to
CC cells requires Mn(II) (By similarity).
CC SUBUNIT: Oligomer with STIM2.
CC SUBCELLULAR LOCATION: Type I membrane protein. Cell surface
CC (Potential).
CC TISSUE SPECIFICITY: Ubiquitously expressed in various human
CC primary cells and tumor cell lines.
CC PTM: Glycosylated.
CC PTM: Phosphorylated.
CC DISEASE: Defects in STIM1 may cause rhabdomyosarcoma and rhabdoid
CC tumors.

CC SIMILARITY: CONTAINS 1 SAM DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U52426; AAC51627.1; -.
DR Genew; HGNC:11386; STIM1.

DR InterPro; IPR001660; SAM.
DR SMART; SM00454; SAM; 1.
DR PROSITE; PS0105; SAM_DOMAIN; 1.
KW Cell adhesion; Anti-oncogene; Transmembrane; Coiled coil; Signal;
KW Glycoprotein; Phosphorylation.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 685 STROMAL INTERACTION MOLECULE 1.
FT DOMAIN 23 213 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 214 234 POTENTIAL.
FT DOMAIN 235 685 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 132 200 SAM.
FT DOMAIN 238 343 COILED COIL (POTENTIAL).
FT DOMAIN 270 336 GLU-RICH.
FT DOMAIN 362 390 COILED COIL (POTENTIAL).
FT DOMAIN 600 629 PRO/SER-RICH.
FT DOMAIN 672 685 LYS-RICH.
FT CARBOHYD 131 131 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 171 171 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 685 AA; 77492 MW; 0AB512CA8D68A7A7 CRC64;

Query Match 11.7%; Score 65.5; DB 1; Length 685;
Best Local Similarity 24.3%; Pred. No. 24;
Matches 25; Conservative 11; Mismatches 38; Indels 29; Gaps 3;
QY 7 EETLWEMESHRRIVRCIPSR-LPFLYLRQAKVLCQDEEVLHSPRLTNSAMRA----- 61
DB 545 DEALNMTSNRRRLIEGVHPGSLV-----EKLDPSPALAKKALLALNHGL 590
QY 62 --GHLLDLLKTRCKNGAATAFLESLEKFNPDVYTLVTGLQPDVD 102
DB 591 DKAHSLMELSPAPPGSGPHLDSRSRSHSPS-----SPDPD 625

RESULT 15
CAML_MOUSE
ID CAML_MOUSE STANDARD; PRT; 1260 AA.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: January 22, 2003, 08:49:36 ; Search time 16.6828 seconds
(without alignments)
1321.544 Million cell updates/sec

Title: us-09-767-215-2_copy_10_116
Perfect score: 560
Sequence: 1 ALTALDEETLWEMESHRR.....PDVYTLVTGLQPDVDFSNFS 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 205047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	270	48.2	1171	4 Q8TES3	Q8tes3 homo sapien
2	75	13.4	467	11 Q99PQ2	Q99pq2 mus musculus
3	73	13.0	210	12 Q66677	Q66677 equine herp
4	73	13.0	265	16 Q97EL5	Q97el5 clostridium
5	73	13.0	311	12 Q9YJN5	Q9yjn5 equine herp
6	71	12.7	265	5 Q61218	Q61218 caenorhabdi
7	70.5	12.6	845	5 Q9V466	Q9v466 drosophila
8	69.5	12.4	522	16 Q9WZD1	Q9wzd1 thermotoga
9	69	12.3	979	5 Q95ZC3	Q95zc3 leishmania
10	68	12.1	221	4 Q9BTQ5	Q9btq5 homo sapien
11	68	12.1	256	4 Q96H24	Q96h24 homo sapien
12	68	12.1	258	4 Q96ET5	Q96et5 homo sapien
13	68	12.1	483	11 Q8VDX5	Q8vdx5 mus musculus
14	68	12.1	519	4 Q8TDN1	Q8tdn1 homo sapien
15	68	12.1	732	13 Q9DDU9	Q9ddu9 xenopus lae
16	67.5	12.1	270	3 Q9P3S8	Q9p3s8 neurospora

17	67.5	12.1	285	12 Q67683	Q67683 groundnut r
18	67.5	12.1	315	5 Q9NA19	Q9nai9 caenorhabdi
19	67.5	12.1	417	5 Q9Y1U6	Q9yliu6 pristonchu
20	67.5	12.1	460	16 Q9CDA4	Q9cda4 mycobacteri
21	67.5	12.1	1009	5 Q8SWY2	Q8swy2 drosophila
22	67.5	12.1	1159	5 Q9NCP8	Q9ncp8 drosophila
23	67.5	12.1	2443	5 Q9VSA2	Q9vsa2 drosophila
24	67	12.0	355	17 Q97AJ9	Q97aj9 thermoplasm
25	67	12.0	374	2 Q93AL8	Q93al8 clostridium
26	67	12.0	431	10 Q9SST8	Q9sst8 icomoea bat
27	67	12.0	1446	10 Q9SUF4	Q9sut4 arabidopsis
28	66.5	11.9	267	12 Q9IE29	Q9ie29 potato viru
29	66.5	11.9	301	12 Q9IHT9	Q9iht9 potato viru
30	66.5	11.9	301	12 Q9IHT8	Q9iht8 potato viru
31	66.5	11.9	301	12 Q9IHT2	Q9iht2 potato viru
32	66.5	11.9	521	16 Q8ZEG5	Q8zeg5 versinia pe
33	66.5	11.9	835	12 Q9E6E6	Q9e6f6 potato viru
34	66.5	11.9	837	12 Q9E6E7	Q9e6f7 potato viru
35	66.5	11.9	3061	12 Q85265	Q85265 potato viru
36	66	11.8	477	4 Q9Y577	Q9y577 homo sapien
37	66	11.8	690	11 Q8R2X3	Q8r2x3 mus musculu
38	65.5	11.7	496	16 Q9KCH4	Q9kch4 bacillus ha
39	65.5	11.7	1259	11 Q9QY38	Q9qy38 mus musculu
40	65	11.6	267	10 Q9FMC1	Q9fmc1 arabidopsis
41	65	11.6	337	17 Q8U341	Q8u341 pyrococcus
42	65	11.6	676	5 Q9W1Q6	Q9w1q6 drosophila
43	65	11.6	768	2 Q9S434	Q9s434 myxococcus
44	64.5	11.5	259	2 Q69441	Q69441 legionella
45	64.5	11.5	320	17 Q30119	Q30119 archaeoglob

ALIGNMENTS

RESULT 1

Q8TES3 PRELIMINARY; PRT; 1171 AA.
AC Q8TES3;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE FLJ00120 protein (Fragment).
GN FLJ00120.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA Ohara O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human spleen."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074049; BAB84875.1; -
FT NON_TER 1
SQ SEQUENCE 1171 AA; 134966 MW; FA567ABBC8A703FF CRC64;

Query Match 48.2%; Score 270; DB 4; Length 1171;
Best Local Similarity 51.5%; Pred. No. 2.9e-22;
Matches 51; Conservative 22; Mismatches 26; Indels 0; Gaps 0;

Qy	6	DDETLWEMESHRRHVRICPSRLTPYLROAKVLCQDDEEVLHSPRLTNSAMRAGHLL	65
Db	35	EDALWENVECNHMLUSRYINPAKLTPYLROCKVDEQDDEVLNAPMLPSKINRAGRL	94
Qy	66	DLIKTRGKNGAIFLESLSKFHNPDVYTLVTGLQPDVDFNS	104
Db	95	DILHTKGQGVVVFLESLEFYYPELYKLYTGKPTRRFS	133

RESULT 2

Q99PQ2

```

ID Q99PQ2 PRELIMINARY; PRT; 467 AA.
AC Q99PQ2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tripartite motif protein TRIM1.
GN TRIM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21231161; PubMed=11331580;
RA Raymond A., Meroni G., Fantozzi A., Merla G., Cairo S., Luzi L.,
RA Riganelli D., Zanaria E., Messali S., Cainarca S., Guffanti A.,
RA Minucci S., Pelicci P.G., Bailly A.,
RT "The tripartite motif family identifies cell compartments."
RL EMBO J. 20:2140-2151(2001).
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL: AF220124; AAG53497.1; -.
DR HSSP: P15919; LRMD.
DR MGD: MGI:2137355; Trim11.
DR InterPro: IPR001870; Gamma_carboxylase.
DR InterPro: IPR003878; SPRY_domain.
DR InterPro: IPR003877; SPRY_receptor.
DR InterPro: IPR000315; Znf_Box.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF00622; SPRY; 1.
DR Pfam: PF00643; zf-B_box; 1.
DR Pfam: PF00097; zf-C3HC4; 1.
DR SMART: SM00336; BBOX; 1.
DR SMART: SM00184; RING; 1.
DR SMART: SM00449; SPRY; 1.
DR PROSITE: PS00518; ZF_RING_1; 1.
DR Zinc-finger.
SQ SEQUENCE 467 AA; 52579 MW; 82B7CF68807E9DA8 CRC64;

Query Match 13.4%; Score 75; DB 11; Length 467;
Best Local Similarity 32.2%; Pred. No. 3.4;
Matches 29; Conservative 17; Mismatches 30; Indels 14; Gaps 5;

QY 7 EFT--LWE-MWESHRIHVRCICPSRLTPYL--RQAKVLCQDEEVLHSPRLTNSAMRA 61
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 EETCALQWQVSEQRQVNLGEF--ERLRLLAEQQLKLEEELEVLPRREGAARL 217
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 GH-----LIDLKTRGNKGAIAFLSLK 84
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 218 GQOSTQALALISELSRQCPALGLLQDIK 247
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
Q66677 PRELIMINARY; PRT; 210 AA.
AC Q66677;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE ORF E10
OS Equine herpesvirus type 2 (strain 86/87) (EHV-2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBI_TaxID=82831;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95302501; PubMed=7783207;
RA Telford E.A., Watson M.S., Aird H.C., Perry J., Davison A.J.;
RT "The DNA sequence of equine herpesvirus 2.";
RL J. Mol. Biol. 249:520-528(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Telford E.A.R.;
RL Submitted (Feb-1995) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL; U20824; AAC13865.1; -.
DR InterPro: IPR001315; CARD.
DR Pfam: PF00619; CARD; 1.
DR SMART: SM00114; CARD; 1.
DR PROSITE: PS50209; CARD; 1.
SQ SEQUENCE 210 AA; 22943 MW; 136D639DD8BC2D40 CRC64;

Query Match 13.0%; Score 73; DB 12; Length 210;
Best Local Similarity 29.5%; Pred. No. 2.3;
Matches 28; Conservative 13; Mismatches 24; Indels 30; Gaps 5;

QY 5 LDEETLWEM-----MESHRIHVRCICPSRLTPYLQAKVLCQDEEVL 49
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16 LTEEDIWVERLCLEELRVLLVSLKSHK-----LDHLRAKKILSREDAEV- 63
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 50 HSPRLTNSAMRAGHLLDLKTRGKNGAIAFLSLK 84
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 -SSRAT-SRSRAGLLVDMCQDHPR-GFQCLKESCK 95
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
Q37EL5 PRELIMINARY; PRT; 265 AA.
AC Q37EL5;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hydroxyethylthiazole kinase Thim/Thik (FS!).
CAC3095.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007806; AAK81035.1; -.
DR InterPro: IPR004399; HMP-P_kinase.
DR TIGRFAMs: TIGR00097; HMP-P_kinase; 1.
KW Kinase; Complete proteome.
SQ SEQUENCE 265 AA; 28539 MW; 327DA3CEAE6D06EB CRC64;

Query Match 13.0%; Score 73; DB 16; Length 265;
Best Local Similarity 23.8%; Pred. No. 3;
Matches 20; Conservative 23; Mismatches 29; Indels 12; Gaps 4;

QY 20 RIVRCICP--SRTPYLQAKVLCQ---DEEVLHSPRLTNS-----AMRAGHLLD-L 67
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 KLITCLPLGLTITPNPEAEVLCGFKIKSEQDMVRAAKKISKINVGILKGHLVDNA 184
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 68 LKTRGKNGAIAFLSLKFNPDVY 91
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 TDLFYKNGYIEWFKLEKIDNPNT 208
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
Q9YJN5 PRELIMINARY; PRT; 311 AA.
AC Q9YJN5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Hypothetical 32.6 kDa protein.
GN E10 OR CLAP.
OS Equine herpesvirus 2.

```

```
OC Viruses; dsDNA viruses, no RNA stage: Herpesviridae;
OC Gammapesvirinae.
OX NCBI_TaxID=133899;
RN [1]
RX MEDLINE=99142601; PubMed=9989495;
RA Willis T.G., Jadavai D.M., Du M.Q., Peng H., Perry A.R.,
RA Abdul-Rauf M., Price H., Karan L., Majekodunmi O., Wlodarska I.,
RA Pan L., Crook T., Hamoudi R., Isaacson P., Dyer M.J.;
RT "Bcl10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and
RT mutated in multiple tumor types.";
RL Cell 96:35-45(1999).
RN [2]
RN SEQUENCE FROM N.A.
RP Thome M., Martinon F., Hofmann K., Rubio V., Steiner V., Schneider P.,
RA "Equine herpesvirus-2 E10, but not its cellular homologue, activates
RA NF-kB transcription factor and c-jun N-terminal kinase.";
RL J. Biol. Chem. 274:17946-17954(1999).
RN [3]
RX MEDLINE=99292766; PubMed=10364242;
RA Srinivasula S.M., Ahmad M., Lin J.-H., Poyet J.-L.,
RA Fernandes-Alnemri T., Tsichlis P.N., Alnemri E.S.;
RT "CUAP, a novel caspase recruitment domain-containing protein in the
RT tumor necrosis factor receptor pathway, regulates NF-kappaB activation
RT and apoptosis.";
RL J. Biol. Chem. 274:17946-17954(1999).
DR EMBL; AJ006410; CAA07016.1; -
DR EMBL; AF100340; AAD16430.1; -
DR EMBL; AF134394; AAD39146.1; -
DR InterPro; IPR001315; CARD.
DR Pfam; PF00619; CARD; 1.
DR SMART; SM00114; CARD; 1.
DR PROSITE; PS02029; CARD; 1.
KW Hypothetical protein.
SQ SEQUENCE 311 AA; 32602 MW; 04EC68C7353D723B CRC64;

Query Match 13.0%; Score 73; DB 12; Length 311;
Best Local Similarity 29.5%; Pred. No. 3.6;
Matches 28; Conservative 13; Mismatches 24; Indels 30; Gaps 5;

QY 5 LDEPTLWEM-----MESHRHVRICPSRLTPYLQAKVLCOLDEEVL 49
| | | | | : | | | | | : | | | | | : | | | | |
| | | | | : | | | | | : | | | | | : | | | | |
Db 16 LFEEDIDVVERLCBELRLVLLSHLKS HKH-----LDHLRAKKILSRDEAEV- 63
| | | | | : | | | | | : | | | | | : | | | | |
QY 50 HSPRLTNSAMRAGHLDDLKTRGKNGAIAFLSLK 84
| | | | | : | | | | | : | | | | | : | | | | |
Db 64 -SSRAT-SRSRAGLLVDMCQDHPR-GFOCLRESCK 95
| | | | | : | | | | | : | | | | | : | | | | |

RESULT 6
O61218 PRELIMINARY; PRT: 265 AA.
ID O61218
AC O61218
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE K03H6.2 protein.
GN Caenorhabditis elegans.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RC STRAIN=BRISTOL N2;
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favallo A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
```

```
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P., III of C.
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RA Wamsley P., Gibson A.;
RT "The sequence of C. elegans cosmid K03H6.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF047658; AAC04418.1; -
DR InterPro; IPR000734; Lipase_3.
DR InterPro; IPR000834; Zn_carboOpept.
DR Pfam; PF01764; Lipase_3; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR PROSITE; PS00120; LIPASE_SER; UNKNOWN_1.
SQ SEQUENCE 265 AA; 30644 MW; E7536832B08BDAFA CRC64;

Query Match 12.7%; Score 71; DB 5; Length 265;
Best Local Similarity 23.9%; Pred. No. 5.1;
Matches 26; Conservative 19; Mismatches 40; Indels 24; Gaps 4;

QY 11 WEMESHRRHVRICPSRLTPYLQAKVLCOLDEEVLHSPRLTNSAMR----- 60
| | | | | : | | | | | : | | | | | : | | | | |
Db 10 WPTMKLSKRLVNCSDPSPILP-LTQCAMITAVDTQKVLVMSFRATNGTQLEEF 68
| | | | | : | | | | | : | | | | | : | | | | |
QY 61 -----AGHL-----DLLKTRGKNGAIAFLSLKFNPDVYTLVTG 96
| | | | | : | | | | | : | | | | | : | | | | |
Db 69 FVAKKAFFDSGVIFEFFVDAYLALWKGLEAEMRLKYRYPDYEVWVTG 117
| | | | | : | | | | | : | | | | | : | | | | |

RESULT 7
Q9V466 PRELIMINARY; PRT: 845 AA.
ID Q9V466
AC Q9V466;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE BCDNA:LD18761 protein.
GN BCDNA:LD18761 OR CG6743.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Besley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
```

```
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Lilang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster." ;
RT Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Tsang G., Brokstein P., Frise E., Harvey D., Evans-Holm M.,
RA Lewis S.E., Suh C., Rubin G.M.,
RT "Full Length Drosophila melanogaster cDNA sequence." ;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003629; AAF53002.1; -
DR EMBL; AF160938; AAD46878.1; -
DR FlyBase: FBgn0027868; BcDNA: LD18761.
SQ SEQUENCE 845 AA; 97381 MW; 0A75B807A6AFD6B6 CRC64;

Query Match 12.6%; Score 70.5; DB 5; Length 845;
Best Local Similarity 30.7%; Pred. NO. 22;
Matches 31; Conservative 13; Mismatches 36; Indels 21; Gaps 5;

QY 2 LFLADEETLWE-----WMESH-----HIVRCICPGRSLTP--YLRQAKVLCQDEEV 48
DB 201 LHAHDMRAWENTLQLENLQAAGKGHKIV-----FLDLPADPVREKRLPHALDEEDN 255

QY 49 LHSPLTNSAMRAGHL---LDLLKTRGKNGATAFLESKFKH 86
DB 256 LRLSRAIFELIRAGRVDDGLKCKHFGQTWRAAILLEGWRLH 296

RESULT 8
Q9WZD1 PRELIMINARY; PRT; 522 AA.
AC Q9WZD1;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein TM0663.
GN TM0663.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima." ;
RL Nature 399:323-329(1999).
```

```
DR EMBL; AE001739; AAD35747.1; -.
DR TIGR; TM0663; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 522 AA; 60668 MW; D7A5E16AA88F79CC CRC64;

Query Match 12.4%; Score 69.5; DB 16; Length 522;
Best Local Similarity 28.4%; Pred. NO. 17;
Matches 31; Conservative 16; Mismatches 41; Indels 21; Gaps 7;

QY 8 ETLEWMESHRRHIVRCICPGRSLTPYLRQAKVLCQDEEVLSHSPRLTN-----SAMRAG 62
DB 348 EPLWKLTDDEER-SFVRVYHGKVTYINLEKDALLRAREEGILLSPERINLHPKHTNVAG 406

QY 63 -----HLLDLKTRGKNGATAFLESKFKHPDVTYLTGLOPDVDFSNF 106
DB 407 YWDPFHILD-LDERVLNGAV-YIHS-----TSEAYT-----EEQEIDAKRF 445

RESULT 9
Q95ZC3 PRELIMINARY; PRT; 979 AA.
AC Q95ZC3;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Possible hypothetical 110.9 kDa protein.
GN Li994.04.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Rieger M.,
RA Ivens A.C., Quail M., Rajandream M.A., Barrell B.G.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome." ;
RL Genome Res. 8:135-145(1998).
DR EMBL; AL596274; CAC44741.1; -.
SQ SEQUENCE 979 AA; 102709 MW; D43039D7F172DE02 CRC64;

Query Match 12.3%; Score 69; DB 5; Length 979;
Best Local Similarity 34.8%; Pred. NO. 40;
Matches 31; Conservative 7; Mismatches 19; Indels 32; Gaps 6;

QY 38 KVLQQLDEEV-LHSPRL---TNSAMRAGHLDDLKT-----RGN 74
DB 187 KVAASDTVLEVLPHAPRLRVPTNLPLEQTLLDLLATATAVVGGPTTEIVKREMGCRG-N 245

QY 75 GATAFLESKFKHP-----DVTYLTG 96
DB 246 PAFALFGE-KFNHPCMLYRWRLYSLLOQ 273

RESULT 10
Q9BTQ5 PRELIMINARY; PRT; 221 AA.
AC Q9BTQ5;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 24.5 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003520; AAH03520.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 1.
DR SMART; SM00320; WD40; 1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 221 AA; 24472 MW; 39CCF8979C10DB53 CRC64;

Query Match 12.1%; Score 68; DB 4; Length 221;
Best Local Similarity 28.7%; Pred. No. 9.2;
Matches 29; Conservative 11; Mismatches 45; Indels 16; Gaps 4;

QY 6 DEE--TLWMESHRRHVRICPCPSRLTPYLKQAKV-----LCQDEEEVLHSP---- 52
   ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 97 DEECTCVVEFGQAGVPVPSPRCSRLRLTHYIEANVGRGYIKELCFSPDGRMISSPHGYG 156
   ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 53 -RLTNSAMRAGHLDLLKTRKNGAIAFLESKLFNPDVYT 92
   ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 157 IRLGLFDKQCSLVCLPKA--SPLRVIRSLYSHNDVLT 195
   ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
Q96H24
ID Q96H24 PRELIMINARY; PRT; 256 AA.
AC Q96H24;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Unknown (protein for MGC:4558).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008969; AAH08969.1; -.
DR InterPro; IPR003131; K_tetra.
DR Pfam; PF02214; K_tetra; 1.
SQ SEQUENCE 256 AA; 29333 MW; 5A48EFC68637F8B8 CRC64;

Query Match 12.1%; Score 68; DB 4; Length 256;
Best Local Similarity 25.2%; Pred. No. 11;
Matches 28; Conservative 20; Mismatches 49; Indels 14; Gaps 4;

QY 1 ALTALDEETLWMESHRRHVRICPCPSRLTPYLKQAKVLCQDEEEVLHSPRLT---N 56
   ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 140 ALSFQELAYWGIEAH---LERC-CLRKLLRKLLEELAKLHREDVLROQRETRRPAS 195
   ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 57 SAMRAGHLDLLKTRGN-----GAIAFLESKLFNPDVYTLVGLQPDV 101
   ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 196 HSSRWGLCMRLRENVPNSQGLPGKVFACLSILEVATTAVSLCVSTMPDL 246
   ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
Q96ET5
ID Q96ET5 PRELIMINARY; PRT; 258 AA.
AC Q96ET5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Hypothetical 28.4 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC TISSUE=OVARY;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 WD REPEAT (TRP-ASP DOMAIN).
DR EMBL; BC011959; AAH11959.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 258 AA; 28385 MW; F50D287B1FABEE9E4 CRC64;

Query Match 12.1%; Score 68; DB 4; Length 258;
Best Local Similarity 28.7%; Pred. No. 11;
Matches 29; Conservative 11; Mismatches 45; Indels 16; Gaps 4;

QY 6 DEE--TLWMESHRRHVRICPCPSRLTPYLKQAKV-----LCQDEEEVLHSP---- 52
   ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 134 DEECTCVVEFGQAGVPVPSPRCSRLRLTHYIEANVGRGYIKELCFSPDGRMISSPHGYG 193
   ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 53 -RLTNSAMRAGHLDLLKTRKNGAIAFLESKLFNPDVYT 92
   ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 194 IRLGLFDKQCSLVCLPKA--SPLRVIRSLYSHNDVLT 232
   ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
Q8VDX5
ID Q8VDX5 PRELIMINARY; PRT; 483 AA.
AC Q8VDX5;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to tripartite motif protein 11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC020102; AAH20102.1; -.
DR InterPro; IPR001870; Gamma_carboxylase.
DR InterPro; IPR003878; SPRY_domain.
DR InterPro; IPR003877; SPRY_receptor.
DR InterPro; IPR000315; Znf_Box.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00622; SPRY; 1.
DR Pfam; PF00643; zf-B_box; 1.
DR Pfam; PF00097; zf-C3HC4; 1.
DR PRINTS; PR01406; BBOXZNFINGER.
DR SMART; SM00336; BBOX; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00449; SPRY; 1.
DR PROSITE; PS00119; ZF_BOX; 1.
DR PROSITE; PS00518; ZF_RING_1; UNKNOWN_1.
DR PROSITE; PS50089; ZF_RING_2; 1.
SQ SEQUENCE 483 AA; 54313 MW; 51BDFDCD314046FC CRC64;

Query Match 12.1%; Score 68; DB 11; Length 483;
Best Local Similarity 32.2%; Pred. No. 23;
Matches 28; Conservative 16; Mismatches 29; Indels 14; Gaps 5;

QY 7 EET--LWE-MESHRRHVRICPCPSRLTPYL--RQAKVLCQDEEEVLHSPRLTNSAMRA 61
   ||| ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 160 EETCALWQKMWESQRQNVLF--ERLRLLAEERQQLQKLEEELEVLPRLREGAARL 217
   ||| ||| :|:| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 62 GH-----LLDLLKTRKNGAIAFLE 81
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 218 GQOSTQLAALISELESRCQLPALGLLQ 244
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
Q8TDNI
ID Q8TDNI PRELIMINARY; PRT; 519 AA.
```

```
AC Q8TDN1:
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Voltage-gated potassium channel Kv6.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ottoschytch N., van Hoorick D., Raes A.L., Snijders D.J.;
RT "Identification, cloning and functional analysis of 3 new human K+
RT channel subunits.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF348984; AAL83911.1; -.
KW Ionic channel.
SQ SEQUENCE 519 AA; 58979 MW; E5BBA354931AB0A4 CRC64;

Query Match 12.1%; Score 68; DB 4; Length 519;
Best Local Similarity 25.2%; Pred. No. 25;
Matches 28; Conservative 20; Mismatches 49; Indels 14; Gaps 4;

QY 1 ALTALDEETLWEMESHRRHVRICPSRLTPYLROAKVLCOLDEEVLHSPRLT-----N 56
Db 140 ALSFQELAYWGIEAH---LERC-CLRLKLLKLELELAKLHREDVLRQRETERPAS 195

QY 57 SAMRAGHLDDLKTRGN-----GAIAFLESKFHNPDPVYTLVTGLQPDV 101
Db 196 HSSRWGLCHNRLREMVENPQSLPGKVFACLSILFVATTAVSLCVSTMPDL 246

RESULT 15
Q9DDU9 PRELIMINARY; PRT; 732 AA.
AC Q9DDU9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Progesterone receptor.
DE XPR-1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20570484; PubMed=11114187;
RA Tian J., Kim S., Heilig E., Ruderman J.V.;
RT "Identification of XPR-1, a progesterone receptor required for Xenopus
RT oocyte activation.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:14358-14363(2000).
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; AF279335; AAG42362.1; -.
DR HSSP; P06536; 1RGD.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR000128; Progester_receptor.
DR InterPro; IPR001723; Stdhrmn_receptor.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF02161; Prog_receptor; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 732 AA; 82022 MW; BF0BA3A14F022199 CRC64;
```

```
Query Match 12.1%; Score 68; DB 13; Length 732;
Best Local Similarity 26.4%; Pred. No. 37;
Matches 29; Conservative 16; Mismatches 35; Indels 30; Gaps 6;

QY 16 SHRHRIVRICP--SRLTPYLROAKVLCOLDEEVL---HSPRLTNSAMRAGHLL----- 65
Db 69 SHLHSWTHFAAPEVSRVTPH-----MCPDSEAKALSIHHTSSLEETSILWAAPTVERKE 122

QY 66 --DLLKTRGKNGAIAFLESKFHNPDP-----VYTLVTGLQPDVDFSN 105
Db 123 PGDSMPLKGNKNS----LEETKEPNSSLTMDSETVTSILPILPDLNLN 168

Search completed: January 22, 2003, 08:56:31
Job time : 20.6828 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 57.4933 Seconds
(without alignments)
683.714 Million cell updates/sec

Title: US-09-767-215-2_COPY_126_420

Perfect score: 1462

Sequence: 1 ECLAGATGSLQELNQEKQ.....TQLRQLQAEPPGVKLQEQART 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1462	100.0	1004	22	AAE07164 Human caspase recr
2	1443	98.7	1139	22	AAE07165 Human predicted ca
3	396.5	27.1	1147	22	AAU01207 Human caspase recr
4	329.5	22.5	1032	23	AAU01206 Human caspase recr
5	329.5	22.5	1032	23	AAU73247 Human plakoglobin
6	283.5	19.4	536	22	AAU01204 Rat caspase recrui
7	274.5	18.8	536	22	AAU01205 Human caspase recr
8	230	15.7	746	22	AAU93822 Human polypeptide,
9	198	13.5	366	22	AAU93617 Human protein sequ
10	197	13.5	2246	22	ABG05850 Novel human diagno

11	196	13.4	612	22	AAU95546	Human protein sequ
12	194	13.3	2101	15	AAU47173	Sequence of the in
13	194	13.3	2101	22	AAU47173	Nuclear mitotic ap
14	194	13.3	2117	22	AAU32040	Novel human secret
15	194	13.3	2192	18	AAU21732	LexA/NumA fusion p
16	194	13.3	2207	22	AAU32041	Novel human secret
17	194	13.3	2272	18	AAU21731	GAL4/HA/NumA fusio
18	193	13.2	2115	21	AAU49937	Human NuMA protein
19	192	13.1	1456	22	ABU58673	Drosophila melanog
20	191.5	13.1	1690	22	ABU61144	Drosophila melanog
21	191.5	13.1	1690	22	ABU61173	Drosophila melanog
22	191	13.1	2101	21	AAU49936	Human NuMA protein
23	190	13.0	2383	23	ABG56631	Human breast speci
24	188.5	12.9	2633	22	ABG06505	Novel human diagno
25	188.5	12.9	2663	22	AAU39097	Human polypeptide
26	188.5	12.9	2688	22	AAU40883	Human polypeptide
27	187.5	12.8	1717	22	ABG20672	Novel human diagno
28	186.5	12.8	2442	21	AAU77575	Human cytoskeletal
29	185	12.7	580	22	AAU42158	Human polypeptide
30	185	12.7	580	22	AAU42159	Human polypeptide
31	185	12.7	690	22	AAU95603	Human protein sequ
32	185	12.7	691	22	AAU40372	Human polypeptide
33	185	12.7	691	22	AAU12179	Human PRO4936 poly
34	185	12.7	691	23	ABU07265	Human APRG polypep
35	185	12.7	711	22	AAU40373	Human polypeptide
36	183	12.5	1694	23	ABU06335	Human GDMPL-1 orth
37	183	12.5	1286	21	AAU43359	Human ORFX ORF3123
38	183	12.5	2053	22	AAU03501	Human protein kina
39	183	12.5	2053	23	AAU24079	Human MDPK protein
40	183	12.5	2054	23	AAE24150	Human kinase (PKIN
41	182	12.4	931	22	ABG23862	Novel human diagno
42	181	12.4	1453	22	AAU39213	Human polypeptide
43	181	12.4	1469	22	AAU39214	Human polypeptide
44	181	12.4	1988	22	AAU40999	Human polypeptide
45	181	12.4	1988	22	AAU41000	Human polypeptide

ALIGNMENTS

RESULT 1

AAE07164

ID AAE07164 standard; Protein; 1004 AA.

XX AC AAE07164;

XX DT 06-NOV-2001 (first entry)

XX DE Human caspase recruitment domain-14 (CARD-14).

XX KW Human; caspase recruitment domain-14; CARD-14; chromosome 17;

XX KW nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy;

XX KW autoimmune disorder; systemic lupus erythematosus; neurological disorder;

XX KW Alzheimer's disease; Parkinson's disease; inflammatory disorder; anaemia;

XX KW haematological disorder; myelodysplastic syndrome; myocardial infarction;

XX KW stroke; immune disorder; Crohn's disease; allergic rhinitis; infection;

XX KW cell signalling disorder; cytostatic; immunosuppressive; nootropic;

XX KW neuroprotective; antiviral; antibacterial.

XX OS Homo sapiens.

XX FH Key

XX FH Modified-site

XX FH Location/Qualifiers

FT FT /note= "CAMP- and cGMP-dependent protein kinase

FT FT phosphorylation site"

FT FT 10..116

FT FT /label= CARD_domain

FT FT Modified-site

FT FT 12..15

FT FT /note= "Casein kinase II phosphorylation site"

FT FT Modified-site

FT FT 18..21

FT FT /note= "Casein kinase II phosphorylation site"

FT FT Modified-site

FT FT 25..27

FT FT /note= "Protein kinase C phosphorylation site"

FT	Modified-site	60..62	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	602..605	/note= "N-glycosylation site"
FT	Modified-site	91..93	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	634..637	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	114..117	/note= "N-glycosylation site"	FT	Modified-site	653..655	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	117..122	/note= "N-myristoylation site"	FT	Modified-site	674..677	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	121..123	/note= "Protein kinase C phosphorylation site"	FT	Domain	676..745	/label= SH3_domain
FT	Domain	126..420	/label= Coiled_Coil_domain	FT	Modified-site	714..719	/note= "N-myristoylation site"
FT	Modified-site	130..135	/note= "N-myristoylation site"	FT	Modified-site	725..727	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	134..137	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	725..728	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	161..166	/note= "N-myristoylation site"	FT	Modified-site	733..738	/note= "N-myristoylation site"
FT	Modified-site	165..168	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	737..740	/note= "N-glycosylation site"
FT	Modified-site	220..227	/note= "Tyrosine kinase phosphorylation site"	FT	Modified-site	759..761	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	221..224	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	760..763	/note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT	Domain	239..325	/label= k-Box_domain	FT	Peptide	785..793	/note= "Peroxisomal targeting signal"
FT	Modified-site	240..243	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	796..799	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	250..252	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	800..805	/note= "N-myristoylation site"
FT	Modified-site	253..256	/note= "Casein kinase II phosphorylation site"	FT	Domain	826..1004	/label= Guanylate_kinase_domain
FT	Modified-site	259..262	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	842..844	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	280..283	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	860..863	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	290..293	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	868..870	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	297..300	/note= "Casein kinase II phosphorylation site"	FT	Region	870..872	/note= "RGD cell attachment sequence"
FT	Modified-site	307..309	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	893..896	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	307..310	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	926..929	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	329..365	/note= "Tyrosine kinase phosphorylation site"	FT	Peptide	941..949	/note= "Peroxisomal targeting signal"
FT	Modified-site	366..368	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	944..947	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	366..369	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	976..979	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	378..381	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	980..985	/note= "N-myristoylation site"
FT	Modified-site	384..386	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	1002..1004	/note= "Protein kinase C phosphorylation site"
FT	Region	385..406	/note= "Leucine zipper pattern"	XX	WO200159065-A2.		
FT	Modified-site	449..452	/note= "Casein kinase II phosphorylation site"	PN	16-AUG-2001.		
FT	Modified-site	463..466	/note= "Casein kinase II phosphorylation site"	PD	22-JAN-2001; 2001WO-US02087.		
FT	Modified-site	463..465	/note= "Protein kinase C phosphorylation site"	PF	09-FEB-2000; 2000US-0181159.		
FT	Modified-site	470..472	/note= "Protein kinase C phosphorylation site"	PR	(MILL-) MILLENNIUM PHARM INC.		
FT	Modified-site	501..504	/note= "Casein kinase II phosphorylation site"	PA	Bertin J;		
FT	Modified-site	511..516	/note= "N-myristoylation site"	XX	WPI; 2001-497073/54.		
FT	Domain	568..660	/label= PDZ_domain	DR	N-PSDB; AAD13447.		
FT	Modified-site	587..592	/note= "N-myristoylation site"	XX	An isolated caspase recruitment domain polypeptide useful for		
FT	Modified-site	589..592		FT			

regulating growth and cell death and useful for the treatment of cancer

Claim 1; Fig 1A-1E; 109pp; English.

The present sequence is human caspase recruitment domain-14 (CARD-14). The CARD-14 gene is located on chromosome 17. The CARD-14 is used for the detection of modulators that modulates the ability of CARD-14 to bind to Bcl-10 and stimulates phosphorylation of Bcl-10 or activation of nuclear factor-kappa B (NF- κ B). The CARD-14 is useful for regulating growth and cell death and useful for the treatment of cancer. It is also useful for the treatment of autoimmune disorders (e.g., systemic lupus erythematosus), neurological disorders (e.g., Alzheimer's and Parkinson's disease, inflammatory disorders), haematological disorders (e.g., anaemia, myelodysplastic syndromes), myocardial infarctions, strokes, immune disorders (e.g., Crohn's disease, allergic rhinitis), cell signalling disorders and certain viral and bacterial infections.

Q	Sequence	1004 AA;
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149		

Query Match 100.0%; Score 1462; DB 22; Length 1004;

Best Local Similarity 100.0%; Pred. No. 6.9e-107;

Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ECLAGISLQEE LNQEKQKEVLLRRCCQLQEHGLGLAETRAEGLHQLEADHSRMKREVS 60

b 126 ECLAGISLQEELNQEKQKEVLLRRCQQLQEHLGLAETRAEGLHQLEADHSRMKREVS 185

61 AHFHEVI.RI.KDEMI.SI.SI.HYSNAI.OEKET.AASRCRSI.OEEI.VI.I.KOELORANMVSSCEI.E 120

DATE	DESCRIPTION	AMOUNT	BALANCE
1/1/20	OPENING BALANCE		100.00
1/15/20	PAYROLL	50.00	150.00
1/20/20	RENT	20.00	130.00
1/25/20	SALES	75.00	205.00
1/30/20	PAYROLL	50.00	255.00
2/5/20	RENT	20.00	235.00
2/10/20	SALES	80.00	315.00
2/15/20	PAYROLL	50.00	365.00
2/20/20	RENT	20.00	345.00
2/25/20	SALES	70.00	415.00
2/28/20	PAYROLL	50.00	465.00
3/5/20	RENT	20.00	445.00
3/10/20	SALES	85.00	530.00
3/15/20	PAYROLL	50.00	580.00
3/20/20	RENT	20.00	560.00
3/25/20	SALES	90.00	650.00
3/30/20	PAYROLL	50.00	700.00
4/5/20	RENT	20.00	680.00
4/10/20	SALES	95.00	775.00
4/15/20	PAYROLL	50.00	825.00
4/20/20	RENT	20.00	805.00
4/25/20	SALES	100.00	905.00
4/30/20	PAYROLL	50.00	955.00
5/5/20	RENT	20.00	935.00
5/10/20	SALES	105.00	1040.00
5/15/20	PAYROLL	50.00	1090.00
5/20/20	RENT	20.00	1070.00
5/25/20	SALES	110.00	1180.00
5/30/20	PAYROLL	50.00	1230.00
6/5/20	RENT	20.00	1210.00
6/10/20	SALES	115.00	1325.00
6/15/20	PAYROLL	50.00	1375.00
6/20/20	RENT	20.00	1355.00
6/25/20	SALES	120.00	1475.00
6/30/20	PAYROLL	50.00	1525.00
7/5/20	RENT	20.00	1505.00
7/10/20	SALES	125.00	1630.00
7/15/20	PAYROLL	50.00	1680.00
7/20/20	RENT	20.00	1660.00
7/25/20	SALES	130.00	1790.00
7/30/20	PAYROLL	50.00	1840.00
8/5/20	RENT	20.00	1820.00
8/10/20	SALES	135.00	1955.00
8/15/20	PAYROLL	50.00	2005.00
8/20/20	RENT	20.00	1985.00
8/25/20	SALES	140.00	2125.00
8/30/20	PAYROLL	50.00	2175.00
9/5/20	RENT	20.00	2155.00
9/10/20	SALES	145.00	2300.00
9/15/20	PAYROLL	50.00	2350.00
9/20/20	RENT	20.00	2330.00
9/25/20	SALES	150.00	2480.00
9/30/20	PAYROLL	50.00	2530.00
10/5/20	RENT	20.00	2510.00
10/10/20	SALES	155.00	2665.00
10/15/20	PAYROLL	50.00	2715.00
10/20/20	RENT	20.00	2695.00
10/25/20	SALES	160.00	2855.00
10/30/20	PAYROLL	50.00	2905.00
11/5/20	RENT	20.00	2885.00
11/10/20	SALES	165.00	3050.00
11/15/20	PAYROLL	50.00	3100.00
11/20/20	RENT	20.00	3080.00
11/25/20	SALES	170.00	3250.00
11/30/20	PAYROLL	50.00	3300.00
12/5/20	RENT	20.00	3280.00
12/10/20	SALES	175.00	3455.00
12/15/20	PAYROLL	50.00	3505.00
12/20/20	RENT	20.00	3485.00
12/25/20	SALES	180.00	3665.00
12/30/20	PAYROLL	50.00	3715.00
1/5/21	RENT	20.00	3695.00
1/10/21	SALES	185.00	3880.00
1/15/21	PAYROLL	50.00	3930.00
1/20/21	RENT	20.00	3910.00
1/25/21	SALES	190.00	4100.00
1/30/21	PAYROLL	50.00	4150.00
2/5/21	RENT	20.00	4130.00

b 186 AHFHEVRLKDEMLSLSHYSNALQEKELAASRCRSLQEELYLLKQELQRANMVSSCELE 245

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Y IZI LQEQSLRTASDQESGDEELNRLKEENERLRSLTFSLAEKDILEQSLDEARGSRQELVERI 180

246 LOEOSLRTASDOESGDEELNRLKEENEKLRSLTFSLAEKDILEOSLDEARGSGROELVER

181 HSLRERAAERQEQWEEKEQTLQFQKSKMACQLYREKVNALQAQVCELQKERDQA

90

D 300 HSLKRAVAERQREQIWEENEQILLQFQKNS

y 241 SARD

b 366 SARSAQREISQSLVEKDSLRRQVFELTDQVCELRTQLRQLQAEPGVLKQEART 420

RESULT 2

AE07165

D AAE07165 standard; Protein; 1139 AA.

X C AAE07165.

2 X
2 'CAT/05000

T 06-NOV-2001 (first entry)

Human predict

Human: caspase recr

nuclear factor- κ B

autoimmune disorder; systemic lupus ery

Alzheimer's dis

haematologist

KW apoptosis; hyperproliferative disorder; autoimmune; neurological;
 KW inflammatory disorder; viral infection; stress-related response.

OS Homo sapiens.

XX	Key	Location/Qualifiers
XX	Domain	6..112 "CARD domain"
XX	Modified-site	7..9 /note= "Protein kinase C phosphorylation site"
XX	Modified-site	7..10 /note= "Casein kinase II phosphorylation site"
XX	Modified-site	100..102 /note= "Protein kinase C phosphorylation site"
XX	Modified-site	100..103 /note= "Casein kinase II phosphorylation site"
XX	Modified-site	105..107 /note= "Protein kinase C phosphorylation site"
XX	Modified-site	106..109 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
XX	Domain	130..431 /note= "Coiled coil domain"
XX	Modified-site	162..165 /note= "Casein kinase II phosphorylation site"
XX	Modified-site	168..171 /note= "Casein kinase II phosphorylation site"
XX	Modified-site	175..183 /note= "Tyrosine kinase phosphorylation site"
XX	Modified-site	182..185 /note= "Casein kinase II phosphorylation site"
XX	Modified-site	189..195 /note= "Tyrosine kinase phosphorylation site"
XX	Modified-site	241..244 /note= "N-glycosylation site"
XX	Modified-site	243..245 /note= "Protein kinase C phosphorylation site"
XX	Modified-site	282..285 /note= "Amidation site"
XX	Modified-site	286..289 /note= "Casein kinase II phosphorylation site"
XX	Modified-site	290..292 /note= "Protein kinase C phosphorylation site"
XX	Modified-site	378..381 /note= "Casein kinase II phosphorylation site"
XX	Modified-site	429..432 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
XX	Modified-site	459..461 /note= "Protein kinase C phosphorylation site"
XX	Modified-site	471..474 /note= "Casein kinase II phosphorylation site"
XX	Modified-site	472..475 /note= "N-glycosylation site"
XX	Modified-site	476..479 /note= "Casein kinase II phosphorylation site"
XX	Modified-site	508..510 /note= "Protein kinase C phosphorylation site"
XX	Modified-site	510..513 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
XX	Modified-site	558..560 /note= "Protein kinase C phosphorylation site"
XX	Modified-site	578..581 /note= "Casein kinase II phosphorylation site"
XX	Modified-site	584..587 /note= "N-glycosylation site"
XX	Modified-site	587..592 /note= "N-myristoylation site"
XX	Modified-site	634..637 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
XX	Domain	635..748

FT	Domain	/note= "PDZ domain"
FT	Modified-site	635..1147 /note= "MAGUK domain"
FT	Modified-site	638..641 /note= "Glycosaminoglycan attachment site"
FT	Modified-site	678..683 /note= "N-myristoylation site"
FT	Modified-site	687..689 /note= "Protein kinase C phosphorylation site"
FT	Modified-site	692..695 /note= "Casein kinase II phosphorylation site"
FT	Modified-site	698..703 /note= "N-myristoylation site"
FT	Modified-site	710..715 /note= "N-myristoylation site"
FT	Modified-site	725..728 /note= "Casein kinase II phosphorylation site"
FT	Modified-site	761..766 /note= "N-myristoylation site"
FT	Modified-site	764..767 /note= "Casein kinase II phosphorylation site"
FT	Domain	766..834 /note= "SH3 domain"
FT	Modified-site	776..779 /note= "N-glycosylation site"
FT	Modified-site	779..782 /note= "Casein kinase II phosphorylation site"
FT	Modified-site	787..789 /note= "Protein kinase C phosphorylation site"
FT	Modified-site	816..819 /note= "Casein kinase II phosphorylation site"
FT	Modified-site	823..828 /note= "N-myristoylation site"
FT	Modified-site	847..850 /note= "Casein kinase II phosphorylation site"
FT	Modified-site	853..858 /note= "N-myristoylation site"
FT	Modified-site	857..859 /note= "Protein kinase C phosphorylation site"
FT	Modified-site	872..875 /note= "Casein kinase II phosphorylation site"
FT	Domain	882..1147 /note= "Guanylate kinase (GUK) domain"
FT	Modified-site	897..900 /note= "Casein kinase II phosphorylation site"
FT	Modified-site	917..922 /note= "N-myristoylation site"
FT	Modified-site	926..929 /note= "Casein kinase II phosphorylation site"
FT	Modified-site	935..937 /note= "Protein kinase C phosphorylation site"
FT	Modified-site	1003..1006 /note= "Casein kinase II phosphorylation site"
FT	Modified-site	1010..1018 /note= "Tyrosine kinase phosphorylation site"
FT	Modified-site	1050..1055 /note= "N-myristoylation site"
FT	Modified-site	1088..1091 /note= "Casein kinase II phosphorylation site"
FT	Modified-site	1120..1123 /note= "Casein kinase II phosphorylation site"
XX	WO200140468-A2.	
XX	07-JUN-2001.	
XX	01-DEC-2000; 2000WO-US32716.	
XX	03-DEC-1999; 99US-0168780.	
XX	18-FEB-2000; 2000US-0507533.	
XX	25-FEB-2000; 2000US-0513904.	
XX	10-OCT-2000; 2000US-0685791.	

FT Modified-site 603..605 /note="Protein kinase C phosphorylation site"
FT 638..641 /note="Glycosaminoglycan attachment site"
FT Modified-site 642..644
FT Modified-site /note="Protein kinase C phosphorylation site"
FT 656..661 /note="N-myristoylation site"
FT Modified-site 681..684
FT /note="Casein kinase II phosphorylation site"
FT Modified-site 690..693
FT /note="Casein kinase II phosphorylation site"
FT Domain 704..772
FT /note="SH3 domain"
FT Modified-site 712..715
FT /note="N-glycosylation site"
FT Modified-site 714..717
FT Modified-site /note="Casein kinase II phosphorylation site"
FT Modified-site 733..739
FT /note="Tyrosine kinase phosphorylation site"
FT Modified-site 748..751
FT /note="Casein kinase II phosphorylation site"
FT Modified-site 754..756
FT /note="Protein kinase C phosphorylation site"
FT Modified-site 754..757
FT /note="Casein kinase II phosphorylation site"
FT Modified-site 761..766
FT /note="N-myristoylation site"
FT Modified-site 782..784
FT /note="Protein kinase C phosphorylation site"
FT Modified-site 809..814
FT /note="N-myristoylation site"
FT 830..1032
FT /note="Guanylate kinase (GUK) domain"
FT Modified-site 830..832
FT Modified-site /note="Protein kinase C phosphorylation site"
FT 868..870
FT /note="Protein kinase C phosphorylation site"
FT Modified-site 869..872
FT /note="Casein kinase II phosphorylation site"
FT Modified-site 882..885
FT /note="Casein kinase II phosphorylation site"
FT Modified-site 893..898
FT /note="N-myristoylation site"
FT Modified-site 915..918
FT /note="Amidation site"
FT Modified-site 947..949
FT /note="Protein kinase C phosphorylation site"
FT Modified-site 981..986
FT /note="N-myristoylation site"
FT Modified-site 1021..1026
FT /note="N-myristoylation site"
FT Modified-site 1022..1024
FT /note="Protein kinase C phosphorylation site"
FT Modified-site 1028..1031
FT /note="Casein kinase II phosphorylation site"

WO200110468-A2.

07-JUN-2001.

01-DEC-2000; 2000WO-US32716.

03-DEC-1999; 99US-0168780.

18-FEB-2000; 2000US-0507533.

25-FEB-2000; 2000US-0513904.

10-OCT-2000; 2000US-0685791.

(MILL-) MILLENNIUM PHARM INC.

Bertin J;

WPI; 2001-367809/38.

DR N-PSDB; AAS05388.
XX Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10,
PT CARD-11, useful as targets for therapy, as immunogens, and in screening
PT and detection assays -
XX
XX Claim 9; Fig 10A-10C; 145pp; English.
XX
CC The present sequence represents novel human caspase recruitment
CC domain, CARD-10. The polynucleotide encoding this sequence was isolated
CC from a human skin cDNA library. Also described are novel human sequences
CC for CARD-9 and CARD-11 (AAU01205, AAU01207) and rat CARD-9 (AAU01204).
CC CARD-9, CARD-10 and CARD-11 interact with Bcl-10 which is thought to
CC activate nuclear factor (NF)-kappaB and apoptosis. The sequences of the
CC invention can be used for treating a disorder associated with abnormal
CC levels of apoptosis by modulating the expression or activity of CARD-9,
CC CARD-10, or CARD-11. They can be used for the treatment of
CC hyperproliferative disorders (e.g. cancer), autoimmune disorders (e.g.
CC systemic lupus erythematosus), neurological disorders (e.g. Crohn's disease),
CC Alzheimer's disease), inflammatory disorders (e.g. HIV). The CARD polypeptide, polynucleotide
CC and viral infection (e.g. HIV). The CARD polypeptide, polynucleotide
CC and an antibody which selectively binds to CARD can be used in screening
CC and detection assays (e.g. chromosomal mapping, tissue typing).
CC predictive medicine (prognostic assays, monitoring clinical trials, and
CC therapy (treatment and prophylaxis). The CARD polypeptide may be used
CC to screen for drugs that bind to and/or modulate it. CARD sequences are
CC potential targets for regulating inflammation, cancer, NF-kappaB
CC signalling, stress-related response and apoptosis in human disease. A
CC host cell containing a polynucleotide encoding CARD can be used to
CC create transgenic animals.
XX
XX Sequence 1032 AA:

Query Match 22.5%; Score 329.5; DB 22; Length 1032;

Best Local Similarity 31.8%; Pred. No. 1.8e-17;

Matches 100; Conservative 50; Mismatches 115; Indels 49; Gaps 6;

QY 10 LOEELNQEKQKEVLLRRCCQQLQEHGLHAETRAEGLHGLEADH-----SRMKREVSAAH 62

Db 136 LMTEVRLRLREARKSQLQREQQQARGVLEEBERAGLEQLRDQQQAQERCQRLREDWEAG 195

QY 63 FHEVLRKDEMLSLSHYNSALQEKELAAASRCRSQEEYLLKQELQANVSSCELELQ 122

Db 196 SUELRKDKENYMIARLAQJSEKNSAVLRSRDLQLAVDQLKLVSR--LEEELCAL--- 250

QY 123 EQSLRTASDOESGDEELNRLKEENE-----KLRS-----LTFSLAE----- 158

Db 251 ---LRRARGPPPGAEEKEKEKEKEPDNDVLSVSELRAENQOLTASLRELOEGLQOEASR 307

QY 159 -----KDLQSLDEARGSRQELVERIHSIRRAVAEROREOYWEKEQTLLOF 208

Db 308 PGAPGSEIRILLDLEHDWREAQDSQELCQKLHAYOGELQWAEELRDQYLQEMEDLRUKH 367

QY 209 OKSKMACOLYREKVNALQAVCELOKEREQDOAYSARDSAQREISQSILVERKDSLRROVFELT 268

Db 368 RFLQKDCDLYKHMTATVLAQLEIEKERDQATQSRDRIQLQYSQSILIEKDYRKQVRGLE 427

QY 269 DQVCELRLTQLRLQ 282

Db 428 AERDELLTTLTSL 441

RESULT 5

AAU73247

ID AAU73247 standard; Protein; 1032 AA.

AC AAU73247;

XX 12-MAR-2002 (first entry)

DT Human plakoglobin interacting protein #3.

DE Human; plakoglobin; cytotstatic; osteopathic; dermatological; cardiant;

XX

KW

KW plakoglobin related disease; skin carcinoma; acantholytic disease;
KW basal cell carcinoma; squamous cell carcinoma; Naxos disease;
KW extramammary Paget's disease; heart disease; skin blistering;
KW subcorneal acantholysis; Grover's disease; Hailey-Hailey's disease;
KW Darier's disease; ectodermal dysplasia; skin fragility syndrome.
XX
OS Homo sapiens.
XX
PN WO200185933-A2.
XX
PD 15-NOV-2001.
XX
XX 02-MAY-2001; 2001WO-EP04872.
PF
XX 09-MAY-2000; 2000EP-0201668.
PR
XX (VLA--) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
PA
XX Van Roy F, Bonne S, Vanlandschoot A;
PI
XX WPI: 2002-062246/08.
XX
DR N-PSDB; AAS98203.
XX
XX New polypeptide, useful for treating skin carcinoma or acantholytic
PT disease such as Grover's and Darier's disease, comprises a protein
PT interacting with human plakoglobin and involved in transduction of
PT plakoglobin related signal to nucleus -
XX
XX Claim 1; Figure 3; 98pp; English.
XX
CC The invention relates to an isolated plakoglobin interacting polypeptide
CC (I). (I) is useful as a medicament and in the manufacture of a
CC medicament for treating plakoglobin related diseases, such as skin
CC carcinoma or an acantholytic disease, and to screen compounds that
CC interfere with the interaction of the polypeptide with plakoglobin
CC The plakoglobin related diseases include basal cell carcinoma, squamous
CC cell carcinoma, extramammary Paget's disease, Naxos disease, heart
CC diseases, skin blistering and acantholytic diseases such as subcorneal
CC acantholysis, Grover's disease, Hailey-Hailey's disease or Darier's
CC disease, and ectodermal dysplasia/skin fragility syndrome. AAU73245-
CC AAU73254 represent novel human plakoglobin interacting protein
CC amino acid sequences and related sequences of the invention.
XX
SQ Sequence 1032 AA;
Query Match 22.5%; Score 329.5; DB 23; Length 1032;
Best Local Similarity 31.8%; Pred. No. 1.8e-17;
Matches 100; Conservative 50; Mismatches 115; Indels 49; Gaps 6;
QY 10 LQELNQRGQREVLRLRCQQQLQEHGLGAEGLHQLQLEADH-----SRMKREVSAAH 62
Db 136 LMTVEVRLREARKSQLOREQQLQARGVLEERAGLEQRLRQOQQAQRCQLREDWEAG 195
QY 63 FHEVLRLKDEMLSLUHYSNALQEKELAAASRCRSQEQELYLLKQELQRANVSSCELELQ 122
Db 196 SLELLRLKDKENTMIAMRLAQLSEKNSAVLSRDLQLAVDQLKLKYSR--LLEECAL---- 250
QY 123 EQSLRTASQESGDEFLNLRKENE-----KLRS-----LTFSLAE----- 158
Db 251 ---LRRARGPPGAEEKEKEKEKEKEDPDNDVLVSELRAENQRLTASLRELQGLQEQEASR 307
QY 159 -----KDLQESLDEARGSRQELVERIHSRLRERAAVAREQYQWEEKQTLQF 208
Db 308 PGAPGSEIRLLDILEHDWEADSRQELCQKLHAGVQELQWAEELRDQYLOEMEDLRLKH 367
QY 209 QKSKMACQLYRKVNALQAVCELAQKRDQAYASARDSAQRETSQSLVKDLSRRQVFELT 268
Db 368 RTLQKDCDLYKHRMATVLAQLIEIKERDQATQSRDRIQLQYSQSLIEKDQYRKQVRGLE 427
QY 269 DQVCELRTQLRQLQ 282
Db 428 AERDELLTTLTSL 441

RESULT 6
AAU01204
ID AAU01204 standard; Protein; 536 AA.
XX
AC AAU01204;
XX
DT 12-SEP-2001 (first entry)
XX
DE Rat caspase recruitment domain, CARD-9 polypeptide.
XX
KW Rat; caspase recruitment domain; CARD-9; Bcl-10; NF-kappaB;
KW apoptosis; hyperproliferative disorder; autoimmune; neurological;
KW inflammatory disorder; viral infection; stress-related response.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT Modified-site 2..5 /note= "Casein kinase II phosphorylation site"
FT Domain 7..98 /note= "CARD domain"
FT Modified-site 12..15 /note= "Casein kinase II phosphorylation site"
FT Modified-site 16..18 /note= "Protein kinase C phosphorylation site"
FT Modified-site 23..26 /note= "Casein kinase II phosphorylation site"
FT Modified-site 92..95 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 95..97 /note= "Protein kinase C phosphorylation site"
FT Modified-site 95..98 /note= "Casein kinase II phosphorylation site"
FT Modified-site 138..140 /note= "Protein kinase C phosphorylation site"
FT Modified-site 138..141 /note= "Protein kinase C phosphorylation site"
FT Domain 140..416 /note= "Casein kinase II phosphorylation site"
FT Modified-site 171..174 /note= "Coiled coil domain"
FT Modified-site 176..183 /note= "Casein kinase II phosphorylation site"
FT Region 197..213 /note= "Tyrosine kinase phosphorylation site"
FT Modified-site 228..231 /note= "Indole-3-glycerol phosphate synthase homology region"
FT Modified-site 231..233 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 267..270 /note= "Protein kinase C phosphorylation site"
FT Region 285..338 /note= "Casein kinase II phosphorylation site"
FT Modified-site 303..305 /note= "Cysteine rich repeat homology region"
FT Modified-site 362..364 /note= "Protein kinase C phosphorylation site"
FT Modified-site 362..365 /note= "Protein kinase C phosphorylation site"
FT Modified-site 374..377 /note= "Casein kinase II phosphorylation site"
FT Modified-site 425..428 /note= "Casein kinase II phosphorylation site"
FT Modified-site 431..433 /note= "Casein kinase II phosphorylation site"
FT Modified-site 451..453 /note= "Protein kinase C phosphorylation site"
FT Modified-site 483..486 /note= "Protein kinase C phosphorylation site"
FT Modified-site /note= "Casein kinase II phosphorylation site"

FT Modified-site 514..516 /note= "Protein kinase C phosphorylation site"

FT Modified-site 523..528 /note= "N-myristoylation site"

FT Modified-site 524..527 /note= "N-glycosylation site"

FT Modified-site 526..529 /note= "Casein kinase II phosphorylation site"

FT Modified-site 531..534 /note= "Casein kinase II phosphorylation site"

XX

PN WO200140468-A2.

XX

PD 07-JUN-2001.

XX

XX 01-DEC-2000; 2000WO-US32716.

PF

XX 03-DEC-1999; 99US-0168780.

XX

PR 18-FEB-2000; 2000US-0507533.

PR

PR 25-FEB-2000; 2000US-0513904.

PR

PR 10-OCT-2000; 2000US-0685791.

XX

XX (MILL-) MILLENNIUM PHARM INC.

PA

PI Bertin J;

XX

XX WPI: 2001-367809/38.

DR

DR N-PSDB; AAS05386.

XX

XX Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10, CARD-11, useful as targets for therapy, as immunogens, and in screening and detection assays -

PT

PT

XX

XX Claim 9; Fig 1A-1B; 145pp; English.

PS

XX The present sequence represents novel rat caspase recruitment domain, CARD-9. The polynucleotide encoding this sequence was isolated from a rat neuronal cDNA library. Also described are novel human sequences for CARD-9, CARD-10 and CARD-11 (AAU01205-AAU01207). CARD-9, CARD-10 and CARD-11 interact with Bcl-10 which is thought to activate nuclear factor (NF)-kappaB and apoptosis. The sequences of the invention can be used for treating a disorder associated with abnormal levels of apoptosis by modulating the expression or activity of CARD-9, CARD-10, or CARD-11. They can be used for the treatment of hyperproliferative disorders (e.g. cancer), autoimmune disorders (e.g. systemic lupus erythematosus), neurological disorders (e.g. Alzheimer's disease), inflammatory disorders (e.g. Crohn's disease), and viral infection (e.g. HIV). The CARD polypeptide, polynucleotide and an antibody which selectively binds to CARD can be used in screening and detection assays (e.g. chromosomal mapping, tissue typing), predictive medicine (prognostic assays, monitoring clinical trials), and therapy (treatment and prophylaxis). The CARD polypeptide may be used to screen for drugs that bind to and/or modulate it. CARD sequences are potential targets for regulating inflammation, cancer, NF-kappaB signalling, stress-related response and apoptosis in human disease. A host cell containing a polynucleotide encoding CARD can be used to create transgenic animals.

XX

SQ Sequence 536 AA;

Query Match 19.4%; Score 283.5; DB 22; Length 536;

Best Local Similarity 29.4%; Pred. No. 3.5e-14;

Matches 91; Conservative 58; Mismatches 123; Indels 37; Gaps 8;

QY 17 EKGQEVLLRRCCQOQLEHL---GLAETRAEGLHOLEADHS-----RMKREVSAAH 62

Db 112 ESGTLQMLTEVMKLOKKVODLTALLSSKDDFKELRVKDSLLRKHOERVORLKECELS 171

QY 63 FHEVRLKDEMLSLSHYSNALQEKELAAASRCRSLQBELLYLLKQELQRANWSSCELELQ 122

Db 172 SAELKRCKDENYDLAMRLAHLSEKGAALMRNRDLQLEVPDLRHSLSMAE--DCKVERK 229

QY 123 EQ-SLRTASDQESGDEELNRLKEENE----KLRLSTLSLAEK-----DILEQSLDE 168

Db 230 HTLKLHAMEQRPQSQELLWDLQQRDLQARVQOELEVSVOEGKLHNRNSPYQVLEEDWRQ 289

QY 169 ARGSRQELVERIHSRLRRVAAARQRYWEEKROTLLQFKSKMACQLYREKYNALQAA 228

Db 290 ALQEHQEQASTIFSLRKDLRQAQEARTRCMEKEKMFELQCLALRKDAKMYKDRTEALQQ 349

QY 229 VCELQKRDQAYSARDSAQREISQSLVEKDSLRQVFEFTDQVCELTQL-----ROLQA 283

Db 350 MEEVSIEDQAMTSREELHAQCAQSFQDKDKLRKQVRELDEKADELQQLFQTESRLAA 409

QY 284 EPPGVLKQE 292

Db 410 E--GRLKQQ 416

RESULT 7

AAU01205

ID AAU01205 standard; Protein; 536 AA.

XX

AC AAU01205;

XX

DT 12-SEP-2001 (first entry)

XX

XX Human caspase recruitment domain, CARD-9 polypeptide.

XX

XX Human; caspase recruitment domain; CARD-9; Bcl-10; NF-kappaB; apoptosis; hyperproliferative disorder; autoimmune; neurological; inflammatory disorder; viral infection; stress-related response.

KW

KW Homo sapiens.

OS

XX

FH Key

FT Modified-site 2..5

FT /note= "Casein kinase II phosphorylation site"

FT Domain

FT 7..98

FT /note= "CARD domain"

FT Modified-site 23..26

FT /note= "Casein kinase II phosphorylation site"

FT Modified-site 92..95

FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"

FT Modified-site 95..97

FT /note= "Protein kinase C phosphorylation site"

FT Modified-site 95..98

FT /note= "Casein kinase II phosphorylation site"

FT Modified-site 138..140

FT /note= "Protein kinase C phosphorylation site"

FT Modified-site 138..141

FT /note= "Casein kinase II phosphorylation site"

FT Domain

FT 140..416

FT /note= "Coiled coil domain"

FT Modified-site 176..183

FT /note= "Tyrosine kinase phosphorylation site"

FT Region

FT 197..213

FT /note= "Indole-3-glycerol phosphate synthase homology region"

FT Modified-site 228..231

FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"

FT Modified-site 231..233

FT /note= "Protein kinase C phosphorylation site"

FT Modified-site 267..270

FT /note= "Casein kinase II phosphorylation site"

FT Region

FT 285..338

FT /note= "Cysteine rich repeat homology region"

FT Modified-site 303..305

FT /note= "Protein kinase C phosphorylation site"

FT Modified-site 333..336

FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"

FT Modified-site 363..366

FT /note= "Casein kinase II phosphorylation site"

Modified-site 425..428
 /note="Casein kinase II phosphorylation site"
 Modified-site 431..433
 /note="Protein kinase C phosphorylation site"
 Modified-site 450..452
 /note="Protein kinase C phosphorylation site"
 Modified-site 453..458
 /note="N-myristoylation site"
 Modified-site 460..462
 /note="Protein kinase C phosphorylation site"
 Modified-site 481..486
 /note="N-myristoylation site"
 Modified-site 483..486
 /note="Casein kinase II phosphorylation site"
 Modified-site 524..527
 /note="N-glycosylation site"
 Modified-site 526..529
 /note="Casein kinase II phosphorylation site"
 Modified-site 527..532
 /note="N-myristoylation site"
 Modified-site 531..534
 /note="Casein kinase II phosphorylation site"
 WO200140468-A2.
 07-JUN-2001.
 01-DEC-2000; 2000WO-US32716.
 03-DEC-1999; 99US-0168780.
 18-FEB-2000; 2000US-0507533.
 25-FEB-2000; 2000US-0513904.
 10-OCT-2000; 2000US-0685791.
 (MILL-) MILLENNIUM PHARM INC.
 Bertin J;
 WPI; 2001-367809/38.
 N-PSDB; AAS05387.
 Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10, CARD-11, useful as targets for therapy, as immunogens, and in screening and detection assays -
 Claim 9; Fig 5A-5B; 145pp; English.
 The present sequence represents novel human caspase recruitment domain, CARD-9. The polynucleotide encoding this sequence was isolated from a human megakaryocyte cDNA library. Also described are novel human sequences for CARD-10 and CARD-11 (AAU01206, AAU01207) and rat CARD-9 (AAU01204). CARD-9, CARD-10 and CARD-11 interact with Bcl-10 which is thought to activate nuclear factor (NF)-kappaB and apoptosis. The sequences of the invention can be used for treating a disorder associated with abnormal levels of apoptosis by modulating the expression or activity of CARD-9, CARD-10, or CARD-11. They can be used for the treatment of hyperproliferative disorders (e.g. cancer), autoimmune disorders (e.g. systemic lupus erythematosus), neurological disorders (e.g. Alzheimer's disease), inflammatory disorders (e.g. Crohn's disease), and viral infection (e.g. HIV). The CARD polypeptide, polynucleotide and an antibody which selectively binds to CARD can be used in screening and detection assays (e.g. chromosomal mapping, tissue typing), predictive medicine (prognostic assays, monitoring clinical trials, and therapy (treatment and prophylaxis). The CARD polypeptide may be used to screen for drugs that bind to and/or modulate it. CARD sequences are potential targets for regulating inflammation, cancer, NF-kappaB signaling, stress-related response and apoptosis in human disease. A host cell containing a polynucleotide encoding CARD can be used to create transgenic animals.

Sequence 536 AA;

Query Match 18.8%; Score 274.5; DB 22; Length 536;

Best Local Similarity 28.4%; Pred. No. 1.8e-13;
 Matches 87; Conservative 61; Mismatches 125; Indels 33; Gaps 7;
 Qy 17 EKGQEVLLRRCCQQLQEH-----GLAETRAEGLHOLEADHS-----RMKREVSAAH 62
 Db 112 ESGLTOLLMTVMKLOKKVQDLTALLSSKDDDFIKELRVKDSLLRKHOERVQRLKECEAG 171
 Qy 63 PHEVLRKDEMLSLSHYSNALQEKELAAASRCRSLOEBLYLLKQELQRANMVSSCELELQ 122
 Db 172 SRELKRCKEENYDLAMRLAHQSEKGAALMRNRLQLEIDQLKHSMLKAE--DDCKVERK 229
 Qy 123 EQ-SLRTASDOESGDEELNRLKEE---NEKLSLTFSLAEK-----DILQSLDE 168
 Db 230 HTLKLRLHAMEQRPQOELLWELOQEKALQARVQLEASVQEGKLDRSSPYIQVLEEDNRQ 289
 Qy 169 ARGSRQELVERIHSRLRERAVAAERQEQWEEKEOTLLQFQKSMACOLYREKVNALQAO 228
 Db 290 ALRDHQEQANTIFSLRKDLROGEARRLRCEKEKEMFLOCLALRKDSMKYKDRIEAILLQ 349
 Qy 229 VCELOKQRDQAYSARDSAQREISQSLVEKDSLRQVVFELTDQVCELRQLROLQAEPPGV 288
 Db 350 MEEVAIERDQAATREELHAQHARGLQEKDALRKQVRELGEKADLEQLQVFOCEAQ---L 406
 Qy 289 LKQEAR 294
 Db 407 LAVEGR 412
 RESULT 8
 AAM93822
 ID AAM93822 standard; Protein; 746 AA.
 XX AC AAM93822;
 XX DT 06-NOV-2001 (first entry)
 XX DE Human polypeptide, SEQ ID NO: 3879.
 XX KW Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX OS Homo sapiens.
 XX PN EP1130094-A2.
 XX PD 05-SEP-2001.
 XX PF 07-JUL-2000; 2000EP-0114089.
 XX PR 08-JUL-1999; 99JP-0194486.
 XX PR 11-JAN-2000; 2000JP-0118774.
 XX PR 02-MAY-2000; 2000JP-0183765.
 XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX DR WPI; 2001-524255/58.
 XX DR N-PSDB; AAK94778.
 XX PT 830 Primers useful for synthesizing full length cDNA clones and their use in genetic manipulation -
 XX PS Claim 8; SEQ ID NO 3879; 1380pp + sequence listing; English.
 CC The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily

CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

SQ		Sequence	746 AA;	
		Query Match	15.7%; Score 230; DB 22; Length 746;	
		Best Local Similarity	41.5%; Pred. No. 8.8e-10;	
Matches	51;	Conservative	25; Mismatches 47; Indels 0; Gaps 0;	
Qy	160	DILEQSDEARGSRQELVERIHSLREAVAAERREQYWEKEOTLLQFOKSKMACOLYR	219 : ::: :: : : : :	
Db	33	DILEHNRREAQDSRGECOKLHAVGSELQWAEELRDQYLQEMLDKHRHTLKQDCDLYK	92 :: : : : : : : : : : : : : : :	
Qy	220	EKNVALAQVCELOKERDAQYSARDSAQRISQLSVFKDSLRRGVFELTDVGCELRTQLR	279 :: : : : : : : : : : : : : : :	
Db	93	HMMATVLQAELLEIEKRDDAIQRSDRTLQOVSQSLIEKKDYRKQVRGLAEAREDELTTLT	152 :: : : : : : : : : : : : : :	
Qy	280	QLQ	282 !:	
Db	153	SLE	155 !!	

RESULT 9	
AAB95617	
ID	AAB95617 standard; Protein; 366 AA.
XX	
XX	
AC	
XX	AAB95617;
XX	
DT	26-JUN-2001 (first entry)
XX	
XX	
DE	Human protein sequence SEQ ID NO:18328.
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX	
XX	
OS	Homo sapiens.
XX	
XX	
PN	EP1074617-A2.
XX	
XX	
PD	07-FEB-2001.

XX primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a

polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB93893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

[illegible]

RESULT 10	
ABG05850	
ID	ABG05850 standard; Protein; 2246 AA.
XX	
AC	ABG05850;
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #5841.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73..
DR	N-PSDB; AAS70037.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX biodiversity -
PS Claim 20; SEQ ID No 36209; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2246 AA;
Query Match 13.5%; Score 197; DB 22; Length 2246;
Best Local Similarity 26.4%; Pred. No. 1.4e-06;
Matches 93; Conservative 47; Mismatches 108; Indels 104; Gaps 14;
Qy 1 ECLAGAGISLOEQLNQGKQEVLLRRCCQQLQEHGLAETRAEGLHQAEDHSMKREVS 60
Db 851 EGLGNLKLCEE-----KNEILQGLKLSLEHLS-----QLQDPPQKGEVL 893
Qy 61 AHFHEVLRLKDEMLSL-----LHYSNALQEKELAAASRCRSLQBELYL----- 103
Db 894 GDVLQLETLKQEAATLANNTQLQARVEMLETERQQQKALLAERGHFEKQQLSSLIT 953
Qy 104 -----LKQLRANMVSSCELEQLQSLRT-----ASDQSGDEELNRLKEE- 145
Db 954 DLQSSISNLSQAKELEQASQAHGARLTAVASLTSELTTLNATTQQQ-DOELAGLKQQA 1012
Qy 146 -----NEKRLSTLFTSLAEKDILEQSLDEARGSRQELVERIHSRL 184
Db 1013 KEKQALQATLQQQEQASQXLRHQVEQLSSLKQK---EQQLKEV-AEKQATRODHA-Q 1067
Qy 185 ERAVAERQREQYWEKEQTLLQFQKSKMACOLYREKYNALQAQVCEQLQKRDQAYSARD 244
Db 1068 QLATAAE-EREASLRERDAALKQLE-----ALEKEKAALKLEI-----LQQQLQVANEARD 1116
Qy 245 SAQRISQSLVEKDSLRRQVFEL-----TDQVCELRITQLRQLQ 282
Db 1117 SAQTSVTAQREKAELSRRKVEELQACVETARQEQHEAQVAQVALEQLRSEQ 1168
RESULT 11
AAB95546
ID AAB95546 standard; Protein: 612 AA.
XX
AC AAB95546;
XX
XX 26-JUN-2001 (first entry)
DT
DE Human protein sequence SEQ ID NO:18167.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
OS
XX

PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
FA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
DR
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 18167; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 612 AA;
Query Match 13.4%; Score 196; DB 22; Length 612;
Best Local Similarity 25.8%; Pred. No. 3.3e-07;
Matches 74; Conservative 62; Mismatches 101; Indels 50; Gaps 12;
Qy 10 LOEELNQGKQEVLLRRCCQQLQEHGLAETRAEGLHQAEDHSMKREVS AHFHEVLRL 69
Db 137 LOEENKILQGRSEELERRVAQLRQ-----IEDLKGDQAKAKETLKKYEGLRL 186
Qy 70 KDEMLSLHYSNALQEKELAAASRCRSLQBELYLKQLRANMVSSCELELQ-EQSLRT 128
Db 187 EEALVHV-----RKEEKEAVSAR-RALENEL-----EAAQGNLSQTTQEQKQLSEKLKE 234
Qy 129 ASDQSGDEELNRLKEEKLRLSLFTSLAEK-----DILEQSLDEARGSRQELVERIHS 182
Db 235 ESEQK---EQLRLKNEMENRHLGKTIKQEKEMADIVEAS-----RTSTLELQNLQDE 287
Qy 183 LRE--RAVAERQREQYWEKEQTLLQFQKSKMACOLYREKYNALQAQVCEQLQKRDQAY 240
Db 288 YKEKNRRELAEQROL-----KEKT-LEAKSRLTAMKWDGCMRLMEELRDYQRAQDEAL 342
Qy 241 SARSAQREISQSLVE-----KDSLRRQVFELTDQVCELRITQLRQ 280
XX

Db 412 VLQETLKQEAATLAANNLTQARVEMLETERGQEQEAKLLAERGHFEEKQQLSSLTIDL 471

QY 104 -----LKOELQRANVSSCELEQESLRT-----ASDQESGDEELNRLKEE--- 145
I::I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I::

Db 472 QSSISNLSQAELEQASQAGHCARLTAQVASUTSELTTLNATIQQQ-DQELAGLQQAQKE 530

QY 146 -----NEKRLSLTSLAEKDILEOSLDEARGSRQELVERIHSRLER 186
::I::I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I::

Db 531 KOAQLAOTLQQOQEQASQGLRHQVEQLSSSLKQK---EQQLKEV-AEKQEAATRDHHA-QQL 585

QY 187 AVAAERQREQYWEKEQTLLOFQKSKMACOLYREKVNALQAQVCELOKRDQAYSARDSA 246
I::I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I::

Db 586 ATAAB-EREASLRERDAALKQLE-----ALEKKAALKLEI-----LQQQLQVANEARDSA 534

QY 247 QREISQSLVEKDSLRQVFEL-----TDQVCELRTOQLRLQ 282
I::I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I::

Db 635 QTSVTAQREKAELSKVVEELQACVETARQEQHEAQAVAELEQLRSEQ 684

RESULT 14

ID AAU32040 standard; Protein; 2117 AA.

XX AC AAU32040;

XX DT 18-DEC-2001 (first entry)

XX DE Novel human secreted protein #2531.

XX KW Human; vaccination; gene therapy; nutritional supplement;

XX KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

XX KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX OS Homo sapiens.

XX PN WO200179449-A2.

XX PD 25-OCT-2001.

XX PF 16-APR-2001; 2001WO-US08656.

XX PR 18-APR-2000; 2000US-0552929.

XX PR 26-JAN-2001; 2001US-0770160.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX PS WPI: 2001-611725/70.

XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic vaccination, testing and therapy -

XX PS Claim 20; Page 547; 765pp; English.

CC The invention relates to novel human secreted polypeptides. The polypeptides and antibodies to the polypeptides are useful for determining the presence of or predisposition to a disease associated with altered levels of polypeptide. The polypeptides are also useful for identifying agents (agonists and antagonists) that bind to them. Cells expressing the proteins are useful for identifying a therapeutic agent for use in treatment of a pathology related to aberrant expression or physiological interactions of the polypeptide. Vectors comprising the nucleic acids encoding the polypeptides and cells genetically engineered to express them are also useful for producing the proteins. The proteins are useful in genetic vaccination, testing and therapy, and can be used as nutritional supplements. They may be used to increase stem cell proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon and/or nerve tissue growth or regeneration; immune suppression and/or stimulation; as anti-inflammatory agents; and in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid sequences of novel human secreted proteins of the invention.

SQ Sequence 2117 AA;

Query Match 13.3%; Score 194; DB 22; Length 2117;
Best Local Similarity 26.6%; Pred. No. 2.2e-06;
Matches 93; Conservative 49; Mismatches 104; Indels 104; Gaps 15;

QY 10 LOEELN---QEK-----GQKEVLLRRCCQLQEHGLAETRAEGLHOLEADHDSRMKREVSAAH 62
I::I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I::

Db 362 LEKELSAALQDKKCLEEKNEILQGLKLSQLEFHL-----QLQDNPPQEKGEVLGD 411

QY 63 PHEVLLRDEMLSLS-----LHYSNALQEKELAAASRCRSLOEELYL----- 103
I::I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I::

Db 412 VLQETLKQEAATLAANNLTQARVEMLETERGQEQEAKLLAERGHFEEKQQLSSLTIDL 471

QY 104 -----LKOELQRANVSSCELEQESLRT-----ASDQESGDEELNRLKEE--- 145
I::I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I::

Db 472 QSSISNLSQAELEQASQAGHCARLTAQVASUTSELTTLNATIQQQ-DQELAGLQQAQKE 530

QY 146 -----NEKRLSLTSLAEKDILEOSLDEARGSRQELVERIHSRLER 186
::I::I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I::

Db 531 KOAQLAOTLQQOQEQASQGLRHQVEQLSSSLKQK---EQQLKEV-AEKQEAATRDHHA-QQL 585

QY 187 AVAAERQREQYWEKEQTLLOFQKSKMACOLYREKVNALQAQVCELOKRDQAYSARDSA 246
I::I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I::

Db 586 ATAAB-EREASLRERDAALKQLE-----ALEKKAALKLEI-----LQQQLQVANEARDSA 634

QY 247 QREISQSLVEKDSLRQVFEL-----TDQVCELRTOQLRLQ 282
I::I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I:: I::I::

Db 635 QTSVTAQREKAELSKVVEELQACVETARQEQHEAQAVAELEQLRSEQ 684

RESULT 15

ID AAW21732 standard; Protein; 2192 AA.

XX AC AAW21732;

XX DT 01-OCT-1997 (first entry)

XX DE LexA/NUMA fusion protein.

XX KW NIP-1; NIP-2; NUMA; nuclear mitotic apparatus; NUMA interacting protein;

XX KW cell division; proliferation; antibody; Ab; detection;

XX KW malignant cell growth.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Domain 1..87

FT Peptide /label= LexA_DNA_binding_domain

FT Protein /label= Polylinker

FT Region 95..2192

FT Region /label= Residues_18-2116_of_NUMA

FT Region 285..1784

FT Region /label= Coiled_coil_region

PN WO9640917-A1.

XX PD 19-DEC-1996.

XX PF 07-JUN-1996; 96WO-US09504.

XX PR 07-JUN-1995; 95US-0478408.

XX PA (UYUA) UNIV YALE.

XX PI McPherson SMG, Snyder MP;

XX DR WPI: 1997-077270/07.

XX DR N-PSDB; AAT77783.

XX PT New nucleic acid encoding nuclear mitotic appts. interacting

PT proteins - useful for modulating cell division and proliferation and
PT in diagnosis

XX PS Claim 15; Page 42-50; 78pp; English.

XX The sequences given in AAW21731-32 represent fusion proteins which
CC contain NuMA (nuclear mitotic apparatus). The fusion proteins were used
CC in the identification of NuMA interacting proteins (NIP's) (see also
CC AAW21729-30). Compounds which interfere with the interaction of NuMA
CC with a known NIP are used to modulate cell division and/or
CC proliferation. Ab, raised conventionally using NIP-1 or -2 as immunogen,
CC are used to detect NIP (or their complexes) and to block their activity
CC for diagnostic or therapeutic use, e.g. to detect defective NuMA or NIP
CC which may be markers for aberrant (including malignant) cell growth
CC (which can also be detected by nucleic acid sequencing). Also where
CC malignancy is related to defects in NuMA or NIP, it can be treated by
CC administration of the appropriate functional protein.

XX SQ Sequence 2192 AA;

Query Match 13.3%; Score 194; DB 18; Length 2192;
Best Local Similarity 26.6%; Pred. No. 2.3e-06;
Matches 93; Conservative 49; Mismatches 104; Indels 104; Gaps 15;
QY 10 LOBELN---QEK---GOKEVLLRCQQLQEHGLAETRAEGLHOLEADHSRMKREVSAAH 62
DB 439 LEKELSAALQDKKCLEEKNEILQGLKSLQLEHLS-----QLQDNPPQEKGEVLGD 488
QY 63 FHEVLKDKEMLSLS-----LHYSNALQEKELAAASRCRSLOEELYL----- 103
DB 489 VLQLETLKQEAATLAANNTOLOARVEMLETERGQOEAKLLAERGHFEKQOOLSLITDL 548
QY 104 -----LKQELQPANMVSSCELELOQSLRT-----ASDOESGDEELNRLKEE--- 145
DB 549 QSSISNLSQAKBELEQASQAHGARLTAQVASLTSELTTLNATIQQQ-DOELAGLKQQAKE 607
QY 146 -----NEKLSLTFSLAEKDILEQSLDEARGSRQELVERIHSLRER 186
DB 608 KQALQATLQQQEQASQGLRHQVEQLSSLKQK---EQOLKEV-AEKQETRDHA-QOL 662
QY 187 AVAARQROYWEEKETLLQFOKSKMACOLYRKVNALQAOVCELOKEDQOAYARDSA 246
DB 663 ATAAE-EREASLRERDAALKQLE-----ALEKEKAAKLEI-----LQOQLOVANEARDSA 711
QY 247 OREISOSLVEKDSLRQVPFL-----TDQVCELRTOLRQLQ 282
DB 712 QTSVTQREKAEKLSRKVEELQACVETARQEQHEAQVQAELEQLRSEQ 761

Search completed: January 22, 2003, 08:52:16
Job time : 63.4933 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	194	13.3	2101	1	US-08-466-390-4	Sequence 4, Appli
2	194	13.3	2101	1	US-08-470-950-4	Sequence 4, Appli
3	194	13.3	2101	1	US-08-467-781-4	Sequence 4, Appli
4	194	13.3	2101	1	US-08-195-487-4	Sequence 4, Appli
5	194	13.3	2101	2	US-08-483-924-4	Sequence 4, Appli
6	194	13.3	2101	4	US-09-452-294-1	Sequence 1, Appli
7	194	13.3	2101	5	PCR-US93-06160-4	Sequence 4, Appli
8	176.5	12.1	976	4	US-09-104-324B-4	Sequence 4, Appli
9	173.5	11.9	2482	1	US-08-328-254-6	Sequence 6, Appli
10	173.5	11.9	3248	1	US-08-353-700-1	Sequence 1, Appli
11	173.5	11.9	3248	5	PCR-US95-16216-1	Sequence 1, Appli
12	171	11.7	1388	2	US-08-685-576-1	Sequence 1, Appli
13	170	11.6	576	2	US-08-533-306A-2	Sequence 2, Appli
14	170	11.6	576	2	US-08-742-923A-2	Sequence 2, Appli
15	170	11.6	816	2	US-08-533-306A-6	Sequence 6, Appli
16	170	11.6	816	2	US-08-742-923A-6	Sequence 6, Appli
17	170	11.6	885	2	US-08-533-306A-4	Sequence 4, Appli
18	170	11.6	885	2	US-08-742-923A-4	Sequence 4, Appli
19	169	11.6	606	4	US-08-477-831C-2	Sequence 2, Appli
20	169	11.6	631	4	US-08-477-831C-11	Sequence 11, Appli
21	167	11.4	712	2	US-08-468-576B-17	Sequence 17, Appli
22	167	11.4	712	2	US-08-468-579B-17	Sequence 17, Appli
23	167	11.4	712	3	US-08-468-577B-17	Sequence 17, Appli
24	164	11.2	1388	2	US-08-685-576-4	Sequence 4, Appli
25	164	11.2	1898	1	US-08-056-200-94	Sequence 94, Appli
26	164	11.2	1898	2	US-08-800-644-94	Sequence 94, Appli
27	162	11.1	1388	4	US-09-572-191-2	Sequence 2, Appli

Db 412 VLQLETLKQEAATLAANNLTQARVEMLETERGQEQEAKLAERGHFEKQKQSLSLITDL 471
QY 104 -----LKQELQANMVSSCELELQEQSLRT-----ASDQSGDEELNRLKEE--- 145
Db 472 QSSISNLSQAKEELEQASQAGHARLTAVASITSELTTLNATIOQO-QDELGLKQOAKE 530
QY 146 -----NEKLSRSLTSLAEKDILEQSLDEARGSRQELVERIHSRLR 186
Db 531 KQQAQLAQTLOQEQEQASQGLRHQVEQLSSSLKQK---EQLKEV-AEKQEAATRDHHA-QOL 585
QY 187 AVAAERQREQYWEKEQTLLQFQKSMACOLYREKVNALQAOVCLOKQERQOAYSARDSA 246
Db 586 ATAABE-EREASLRERDAALKQLE-----ALEKKAALKLEI-----LQOQLQVANEARDSA 634
QY 247 QREISQSLVEKSLRQVPEL-----TDQVCLELRTQLRLQ 282
Db 635 QTSVTQAOQREKAELSRKVELOACVETARQEQHEAQAOQVAELEQLRSEQ 684

RESULT 2

US-08-470-950-4
; Sequence 4, Application US/08470950
; Patent No. 5698439
; GENERAL INFORMATION:
; APPLICANT: TOUTKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/470,950
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-470-950-4

Query Match 13.3%; Score 194; DB 1; Length 2101;
Best Local Similarity 26.6%; Pred. No. 9.6e-08;
Matches 93; Conservative 49; Mismatches 104; Indels 104; Gaps 15;

QY 10 LQELN---QEK-----GQKEVLLRRCCQQLQEHGLAETRAEGLHQAADHSMKREVSAAH 62
Db 362 LEKLSAALQDKKCLEKNEILQKLSQLEHLS-----QLQDNPPQEKGEVLGD 411
QY 63 FHEVLRKDEMLSL-----LHYSNALQEKELAAASRCRSLQBELYL----- 103
Db 412 VLQLETLKQEAATLAANNLTQARVEMLETERGQEQEAKLAERGHFEKQKQSLSLITDL 471

QY 104 -----LKQELQANMVSSCELELQEQSLRT-----ASDQSGDEELNRLKEE--- 145
Db 472 QSSISNLSQAKEELEQASQAGHARLTAVASITSELTTLNATIOQO-QDELGLKQOAKE 530
QY 146 -----NEKLSRSLTSLAEKDILEQSLDEARGSRQELVERIHSRLR 186
Db 531 KQQAQLAQTLOQEQEQASQGLRHQVEQLSSSLKQK---EQLKEV-AEKQEAATRDHHA-QOL 585
QY 187 AVAAERQREQYWEKEQTLLQFQKSMACOLYREKVNALQAOVCLOKQERQOAYSARDSA 246
Db 586 ATAABE-EREASLRERDAALKQLE-----ALEKKAALKLEI-----LQOQLQVANEARDSA 634
QY 247 QREISQSLVEKSLRQVPEL-----TDQVCLELRTQLRLQ 282
Db 635 QTSVTQAOQREKAELSRKVELOACVETARQEQHEAQAOQVAELEQLRSEQ 684

RESULT 3

US-08-467-781-4
; Sequence 4, Application US/08467781
; Patent No. 5780596
; GENERAL INFORMATION:
; APPLICANT: TOUTKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/467,781
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-781-4

Query Match 13.3%; Score 194; DB 1; Length 2101;
Best Local Similarity 26.6%; Pred. No. 9.6e-08;
Matches 93; Conservative 49; Mismatches 104; Indels 104; Gaps 15;

QY 10 LQELN---QEK-----GQKEVLLRRCCQQLQEHGLAETRAEGLHQAADHSMKREVSAAH 62
Db 362 LEKLSAALQDKKCLEKNEILQKLSQLEHLS-----QLQDNPPQEKGEVLGD 411
QY 63 FHEVLRKDEMLSL-----LHYSNALQEKELAAASRCRSLQBELYL----- 103
Db 412 VLQLETLKQEAATLAANNLTQARVEMLETERGQEQEAKLAERGHFEKQKQSLSLITDL 471
QY 104 -----LKQELQANMVSSCELELQEQSLRT-----ASDQSGDEELNRLKEE--- 145
Db 472 QSSISNLSQAKEELEQASQAGHARLTAVASITSELTTLNATIOQO-QDELGLKQOAKE 530

QY 146 -----NEKLSRSLTFLAEKDILQSLDEARGSRQELVERIHSLR 186
DB 531 KOAQLAQTLLQOEQAQSGRLRHQVEQLSSSLKQK---EQQLKEV-AEKQEAATRDHA-QQL 585
QY 187 AVAAERQRYWEKEQTLLQFOKSKMACOLYREKYNALQAOVCCELOKQERDQAYSARDSA 246
DB 586 ATAAE-EREASURERDAALKQLE-----ALEKEKAACLEI-----LQOQLQVANEARDSA 634
QY 247 QREISQSLVEKDSLRQVFELE-----TDQVCELRQLTQLRQLQ 282
DB 635 QTSVTOAQREKAELSRRVLEQACVETARQEQHEAQVAELELQURSEQ 684

RESULT 4
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOUTKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-195-487-4

Query Match 13.3%; Score 194; DB 1; Length 2101;
Best Local Similarity 26.6%; Pred. No. 9.6e-08;
Matches 93; Conservative 49; Mismatches 104; Indels 104; Gaps 15;
QY 10 LOBELN---QEK-----GQKEVLLRRCCQQLQEHGLAETRAEGLHQLEADHSRMKREVSAAH 62
DB 362 LEKELSAALQDKKCLEKEKNEILQGLKLSLEHLS-----LQDNPPQKEGVLGD 411
QY 63 FHEVLKDKEMLSL-----LHYSNALQEKELAAASRCRSLQBELYL----- 103
DB 412 VLQLETLKQEAATLAAANTQLOARVEMLETERGQQEAKLLAERGHFEKQQLSSLTDL 471
QY 104 -----LKQELQRANMVSSCELELOQSRLT-----ASDOESGDEELNRLKEE--- 145
DB 472 QSSISNLSQAKELEQAQSAHGARLTAQVASLTSELTTLNATIQQQ-QDELAKLKQAKE 530

QY 146 -----NEKLSRSLTFLAEKDILQSLDEARGSRQELVERIHSLR 186
DB 531 KOAQLAQTLLQOEQAQSGRLRHQVEQLSSSLKQK---EQQLKEV-AEKQEAATRDHA-QQL 585
QY 187 AVAAERQRYWEKEQTLLQFOKSKMACOLYREKYNALQAOVCCELOKQERDQAYSARDSA 246
DB 586 ATAAE-EREASURERDAALKQLE-----ALEKEKAACLEI-----LQOQLQVANEARDSA 634
QY 247 QREISQSLVEKDSLRQVFELE-----TDQVCELRQLTQLRQLQ 282
DB 635 QTSVTOAQREKAELSRRVLEQACVETARQEQHEAQVAELELQURSEQ 684
RESULT 5
US-08-483-924-4
; Sequence 4, Application US/08483924
; Patent No. 5882876
; GENERAL INFORMATION:
; APPLICANT: TOUTKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,924
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-924-4

Query Match 13.3%; Score 194; DB 2; Length 2101;
Best Local Similarity 26.6%; Pred. No. 9.6e-08;
Matches 93; Conservative 49; Mismatches 104; Indels 104; Gaps 15;
QY 10 LOBELN---QEK-----GQKEVLLRRCCQQLQEHGLAETRAEGLHQLEADHSRMKREVSAAH 62
DB 362 LEKELSAALQDKKCLEKEKNEILQGLKLSLEHLS-----LQDNPPQKEGVLGD 411
QY 63 FHEVLKDKEMLSL-----LHYSNALQEKELAAASRCRSLQBELYL----- 103
DB 412 VLQLETLKQEAATLAAANTQLOARVEMLETERGQQEAKLLAERGHFEKQQLSSLTDL 471
QY 104 -----LKQELQRANMVSSCELELOQSRLT-----ASDOESGDEELNRLKEE--- 145
DB 472 QSSISNLSQAKELEQAQSAHGARLTAQVASLTSELTTLNATIQQQ-QDELAKLKQAKE 530
QY 146 -----NEKLSRSLTFLAEKDILQSLDEARGSRQELVERIHSLR 186

```
Db 531 KQAQLAQTLLQQEQASQGLRHQVEQLSSSLKQ-----EQQLKEV-AEKQEAATRDHHA-QOL 585
QY 187 AVAAERQRYWEEKEQTLQFOKSKMACOLYREKVNALQAQVCELOKEROQAYSARDSA 246
Db 586 ATAEE-EREASLRERDAALKOLE-----ALEKEAKALEI-----LQOOLQVANEARDSA 634
QY 247 QREISQSLVEKDSLRROVFEL-----TDQVCELRTOQLRQL 282
Db 635 QTSVTQAQREKAELSRKVEELQACVETARQEQHEAQVAQVAEELQLRSEQ 684

RESULT 6
US-09-452-294-1
; Sequence 1, Application US/09452294
; Patent No. 6287790
; GENERAL INFORMATION:
; APPLICANT: Lelievre, Sophie
; APPLICANT: Bissell, Mina
; TITLE OF INVENTION: UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED
; TITLE OF INVENTION: THERAPY AND DETECTION OF PROLIFERATIVE AND
; TITLE OF INVENTION: DIFFERENTIATION DISORDERS
; FILE REFERENCE: IB-1454- Sequence Submittal
; Patent No. 6287790
; CURRENT APPLICATION NUMBER: US/09/452,294
; CURRENT FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: 60/110,420
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 2101
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-452-294-1

Query Match 13.3%; Score 194; DB 4; Length 2101;
Best Local Similarity 26.6%; Pred. No. 9.6e-08;
Matches 93; Conservative 49; Mismatches 104; Indels 104; Gaps 15;

QY 10 LQELN---QEK---GQEVLLRRCQQLQEHGLHGLAETRAEGLHOLEADHSMRKREVSAAH 62
Db 362 LEKLSAALQDKCKLEBKNEILQGLKSQLEEHL-----QLQDNPPEKGEVLGD 411
QY 63 FHEVRLKDEMLS-----LHYSNALQEKELAAASRCRSLOEELYL----- 103
Db 412 VLQETLKQEAATLAAANTTQQAQVEMLETERGQQAELKLAERGHFEEKQQLSSLITDL 471
QY 104 -----LKQELQANVSSCELELOEQSLRT-----ASDQSGDEELNRLKEE--- 145
Db 472 QSSISNLQAEELEQASQAHGARLTAQVASTLTTLNATIQQQ-QDELQAGLQQAQKE 530
QY 146 -----NEKLSRSTFSLAEKDIQSLDEARGSRQELVERIHSRLR 186
Db 531 KQAQLAQTLLQQEQASQGLRHQVEQLSSSLKQ---EQQLKEV-AEKQEAATRDHHA-QOL 585
QY 187 AVAAERQRYWEEKEQTLQFOKSKMACOLYREKVNALQAQVCELOKEROQAYSARDSA 246
Db 586 ATAEE-EREASLRERDAALKOLE-----ALEKEAKALEI-----LQOOLQVANEARDSA 634
QY 247 QREISQSLVEKDSLRROVFEL-----TDQVCELRTOQLRQL 282
Db 635 QTSVTQAQREKAELSRKVEELQACVETARQEQHEAQVAQVAEELQLRSEQ 684

RESULT 7
PCT-US93-06160-4
; Sequence 4, Application PC/TUS9306160
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06160
; FILING DATE: 19930621
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESO, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-06160-4

Query Match 13.3%; Score 194; DB 5; Length 2101;
Best Local Similarity 26.6%; Pred. No. 9.6e-08;
Matches 93; Conservative 49; Mismatches 104; Indels 104; Gaps 15;

QY 10 LQELN---QEK---GQEVLLRRCQQLQEHGLHGLAETRAEGLHOLEADHSMRKREVSAAH 62
Db 362 LEKLSAALQDKCKLEBKNEILQGLKSQLEEHL-----QLQDNPPEKGEVLGD 411
QY 63 FHEVRLKDEMLS-----LHYSNALQEKELAAASRCRSLOEELYL----- 103
Db 412 VLQETLKQEAATLAAANTTQQAQVEMLETERGQQAELKLAERGHFEEKQQLSSLITDL 471
QY 104 -----LKQELQANVSSCELELOEQSLRT-----ASDQSGDEELNRLKEE--- 145
Db 472 QSSISNLQAEELEQASQAHGARLTAQVASTLTTLNATIQQQ-QDELQAGLQQAQKE 530
QY 146 -----NEKLSRSTFSLAEKDIQSLDEARGSRQELVERIHSRLR 186
Db 531 KQAQLAQTLLQQEQASQGLRHQVEQLSSSLKQ---EQQLKEV-AEKQEAATRDHHA-QOL 585
QY 187 AVAAERQRYWEEKEQTLQFOKSKMACOLYREKVNALQAQVCELOKEROQAYSARDSA 246
Db 586 ATAEE-EREASLRERDAALKOLE-----ALEKEAKALEI-----LQOOLQVANEARDSA 634
QY 247 QREISQSLVEKDSLRROVFEL-----TDQVCELRTOQLRQL 282
Db 635 QTSVTQAQREKAELSRKVEELQACVETARQEQHEAQVAQVAEELQLRSEQ 684

RESULT 8
US-09-104-324B-4
; Sequence 4, Application US/09104324B
; Patent No. 6232460
; GENERAL INFORMATION:
; APPLICANT: T rec1, Ozlem; Sahin, Ugur; Pfreundschuh, Michael
; TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
; TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample
; TITLE OF INVENTION: No. 6232460mal Cells
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Fulbright & Jaworski LLP
; STREET: 666 Fifth Avenue
; CITY: New York City
```

STATE: New York
ZIP: 10103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/104,324B
FILING DATE: 25-June-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/892,702
FILING DATE: 15-July-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6232460man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5491
TELEPHONE: (212) 318-3000
TELEFAX: (212) 752-5958
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 976 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-104-324B-4

Query Match 12.1%; Score 176.5; DB 4; Length 976;
Best Local Similarity 23.9%; Pred. No. 1.le-06;
Matches 71; Conservative 53; Mismatches 112; Indels 61; Gaps 10;
QY 12 EELNEKQGEVLLRRCCQL-----OEHLGLAETRAEGLHOLEADHSRMKREVS 60
DB 434 BELKKVLEKEITLVENKQFEKIAELKTEQELIGLQARKEVHDLIEIQTAITTSQ 493
QY 61 AHFHEVRLKDEMLSLSHYSNALQEKELAAASCRSLQEEYLLKQELQARANNVSCSELE 120
DB 494 YYSKEVKDLKTEL-----ENEKLNTEL-TSHCNKLSLENKELTQE-----TSDMTLE 540
QY 121 LQEQSLRTASDQSGDEELNRLKEENEKRLSLTFSLAE-----KDILEQSLDSE 168
DB 541 LKNQ-----QEDINNNKQOEERMLKQIENIQETETOLRNELEYVREELKQKRDE 589
QY 169 ARGSRQELVERIHSRLRERAAERQYWEKEOTLLQFQSKMA-----COLYREKVA 224
DB 590 VKCKLDKSENCNNLRKQV-----ENKNYIEELQENKALKKGTAEKQLNLYEIKVVK 645
QY 225 LQAQVCELQERDQAYSARDSAQREISQSLVEKDSLRQVFE---LTDQVCELRQL 278
DB 646 LE---LELESQKQKGEITDVTQKETEDKKISEENLLEEVERAKVIADEAVKLQKEI 699

RESULT 9
US-08-328-254-6
Sequence 6, Application US/08328254
Patent No. 5710022
GENERAL INFORMATION:
APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hwa
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,254
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/141,239
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 1191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2482 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-328-254-6
Query Match 11.9%; Score 173.5; DB 1; Length 2482;
Best Local Similarity 22.4%; Pred. No. 5.9e-06;
Matches 89; Conservative 66; Mismatches 134; Indels 109; Gaps 13;
QY 3 LAGATGSLQEEELNQEKGQKEVLLRRCCQLQOEHLGLAETRAEGL-HOLE-----ADHSRMK 56
DB 1667 LRNSEKLEKRLARLEADEKQKQCVLQQLKSEHHDLLKGRVENLERELEIARTNOEHALE 1726
QY 57 REVSAHFHEVRLKDEMLSLSH-----Y 80
DB 1727 AENSKGEVETLAKITGTMQSLRGLELDVVITRSEKENTNELOKEQERISELEIINSF 1786
QY 81 SNALQEK-----ELAAASRCRSLQEEYLL--KOELQRA---NMVSCE-L 119
DB 1787 ENILQKEQEKVQMKESSTAMEMLQTLQKELNERNVAALHNDQEACKQENLSQVECL 1846
QY 120 ELQEOSLRASDQ-----ESGDEELNRLKEENEKRLSLT 153
DB 1847 ELEKALQGLQDEAKNNYIVLOSSVNLQOEVEDGKQKLEKDEEISRLKNOIQDOEQLV 1906
QY 154 FSLAEKD-----ILEQSLDEARGSRQELVERIHSRLRERAAERQYWEKEQTLLOF 208
DB 1907 SKLSQVEGEHLWKEQNL-ELRNLTVELQKIQVLOSKNASLQDTLEVLQSSYKNLENEL 1965
QY 209 QKSKMACOLYREKVNALQAVCELOKQERDQAYSARDSAQREIS-----QSLVEK- 257
DB 1966 ELTKMDKMSFVEKVKNKMTAKETELQREHHEMAQKTAELQEBELSGEKNRLAGELQLLEEI 2025
QY 258 DSLRQVFEELTDQVCELRQL---RQLOAEPGVLKQE 292
DB 2026 KSKDKLQKELTLENSELKSKSLDCMHKQVKEGKVRREE 2063
RESULT 10
US-08-353-700-1
Sequence 1, Application US/08353700
Patent No. 5599919
GENERAL INFORMATION:
APPLICANT: YEN, TIMOTHY J.
APPLICANT: RATTNER, JEROME B.
TITLE OF INVENTION: NUCLEIC ACID ENCODING A
TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA

us-09-767-

AFFILIATION: Gen, Timothy O.
 APPLICANT: Ratner, Jerome B.
 TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
 Expressed Kinetochore Protein, and Methods of Use
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dann, Dorfman, Herrell and Skillman

	Query Match	11.7%	Score	171;	DB	2;	Length	1388;
	Best Local Similarity	22.1%;	Pred.	No.	4.7e-06;			
	Matches	73;	Conservative	67;	Mismatches	115;	Indels	76; Gaps
								12;
QY	3	LAGAIGSLQEELNQEKGVLLRRQ-----QLQEHGLGAEATRAEGL-----HOLEADHS	53					
Db	644	LOGRISGLEEDVKNKG---ILLAKVELEKROLQGERFTDEKEKNNEIDMTYOLKLVIQQ	699	:	:	:	:	:
QY	54	RMKREVSAAHPFVLRLXDE--MLSLSLHYSNALQVELAASRCRSLOEFELYLLKQELQR	110	:	:	:	:	:
Db	700	SLEGTEETHKATRKLARLDANKKIYESIEAKSEAMKEKKLSEERTLKQKVENLLLAEAK	759	:	:	:	:	:
QY	111	ANMVSSCELELQEQSLR-----TASDOESGDDELN-	140	:	:	:	:	:
Db	760	RCSILDCLQKSQOKINELLKQDVNLDEDVRNLTLKETQTKKLCITONDLMQMTOQQVNT	819	:	:	:	:	:
QY	141	-----RKBEKENELRSUTFSLAEKDI-LQSLEDAGSRGEVLVERIHSLRERAVAAERQ	193	:	:	:	:	:
Db	820	LKWSEKQLKOENNHLLEMKWSLEKQNAELRKERODADGMKQLDQL-----E	867	:	:	:	:	:
QY	194	REQYWEEKOTLTLOFKSKMACQLRYREKVNVALOQCVCQLKERDAQVSARDSAQREISQS	253	:	:	:	:	:
Db	868	AEOFYSTLYKT-QVRELKBCBCKTKLCKELOKKQELQDERDSL-----AAQLIEITLT	920	:	:	:	:	:
QY	254	LVEKDSSLRRQVFELTDQVCFLRTO--LRQLQ	282	:	:	:	:	:

Query Match	11.6%	Score 170;	DB 2;	Length 576;
Best Local Similarity	21.0%;	Pred. No. 1.9e-06;		
Matches	70;	Conservative 73;	Mismatches 108;	Indels 82; Gaps 12;
Qy	11	QEELENOEKGOKEVLLRRCCQQLQOEHLGLAETRAEGLHOLEADHSRKMKEVSAHFHEVLRK	70	
Db	221	RDEONEEK-----RR--QIQRQLHVEYET-----ELEDERNERALAAAAKK--KLE	261	
Qy	71	DEMLSLSHYGNALQEKELAAASRCSLQOEELYLLKQELQ-----RAN	112	
Db	262	GDLDLELQADSAIKGREETKQLRKLOAQMKDFQRELEADARASDEIFATAKENEKRAK	321	
Qy	113	MVSSCELELQOSLRTASDQSGDEELNRLKEE-----NEKURSLTFSLAEKDI	161	
Db	322	SLEADLMQLOLEDAARAKRQADLEKELEAEELASSLSGRNALQDEKRRRLAETIAQ---	378	
Qy	162	LEQSIDAEARGSRQELVERIHSRLRRA-----VAAERQEQYWEKEQTLQLQOFQSKMAC	215	
Db	379	LEEELEEBQGNWEMASDRVRKATQQAQOLSNELAYETSAQKNESARQO--LERQNKELRS	437	
Qy	216	QL-----YREKYNALQAQVCELQKQRDQAYSARDSAQRETSQS-----LVEKD	258	
Db	438	KLHEMEGAVKSKFNSTIAALEAKTAQLEEQVEQEAREREQAATKSLKQDKDKKLLELLOVE	497	
Qy	259	SLRRQVFPETDQ-----VCELRTQLRQLQAE	284	
Db	498	DERKMAEQYKEQAEKGNARVKOLKROLPEAEFEE	530	

RESULT 14
US-08-742-923A-2
: Sequence 2, Application US/08742923A
: Patent No. 5869611
: GENERAL INFORMATION:
: APPLICANT: Liu, Pu
: APPLICANT: Collins, Francis S.
: APPLICANT: Siciliano, Michael J.
: APPLICANT: Claxton, David
: TITLE OF INVENTION: Markers for Detection of Chromosome 16
: TITLE OF INVENTION: Rearrangements
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
: STREET: P.O. Box 828
: CITY: Bloomfield Hills
: STATE: MI
: COUNTRY: USA
: ZIP: 48303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/742,923A
: FILING DATE: No. 5869611ember 1, 1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Smith, DeAnn F.
: REGISTRATION NUMBER: 36683
: REFERENCE/DOCKET NUMBER: 2115-00869DVC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (810) 641-1600
: TELEFAX: (810) 641-0270
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 576 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-742-923A-2

Query Match 11.6%; Score 170; DB 2; Length 576;
Best Local Similarity 21.0%; Pred. No. 1.9e-06;
Matches 70; Conservative 73; Mismatches 108; Indels 82; Gaps 12;
QY 11 QEELNOEKGQKEVLLRRCCQQLQEHGLAETRAEGLHQLEADHSRMRKREVS AHFHEVLR LK 70
Db 221 RDEQNEEK-----RR--QLQRLHEYET-----ELEDERNERALAAAKK--KLE 261
QY 71 DEMLSLSLHYSNALQEKELAAASRCRSLOEELYLKQELQ-----RAN 112
Db 262 GDLKDLQLEQADSAIKGREAIAKQLRKLAQMKDFQLEEDARASRDEIFATAKENEKKA 321
QY 113 MYSSCELELQEQSLRTASDQESGDEELNRLKEE-----NEKLSRSLTFSLAEKDI 161
Db 322 SLEADLMQLQEDLAAARAKQADLEKEELAEELASSLSGRNALQDEKRRLEARIAQ--- 378
QY 162 LEQSLDEARGSRQELVETHSLRERA-----VAAERQEQYWEKEQTLLQFQSKMKAC 215
Db 379 LEELEEEEGNMEAMSDRVKRTAQAEQLSNELATERSTAQNESARQO--LERQNKELRS 437
QY 216 QL-----YREKVNALQAVCELOKQDQAYSARDSAORETSQS-----LVEKD 258
Db 438 KLHEMGAVKSKFKSTIAALEAKIAQLEQVEQEAEREQAATKSLKQDKKLKILLQVE 497
QY 259 SLRRQVFELTDQ-----VCELRITQLRQLQAE 284
Db 498 DERKMAEQYKEQAEKGNARVKQLKRLQLEAEAE 530

RESULT 15
US-08-533-306A-6
: Sequence 6, Application US/08533306A
: Patent No. 5837457
: GENERAL INFORMATION:
: APPLICANT: Liu, Pu
: APPLICANT: Collins, Francis S.
: APPLICANT: Siciliano, Michael J.
: APPLICANT: Claxton, David
: TITLE OF INVENTION: Markers for Detection of Chromosome 16
: TITLE OF INVENTION: Rearrangements
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
: STREET: P.O. Box 828
: CITY: Bloomfield Hills
: STATE: MI
: COUNTRY: USA
: ZIP: 48303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/533,306A
: FILING DATE: September 25, 1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Smith, DeAnn F.
: REGISTRATION NUMBER: 36683
: REFERENCE/DOCKET NUMBER: 2115-00869COB
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (810) 641-1600
: TELEFAX: (810) 641-0270
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 816 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-533-306A-6

Query Match 11.6%; Score 170; DB 2; Length 816;
Best Local Similarity 21.0%; Pred. No. 2.9e-06;
Matches 70; Conservative 73; Mismatches 108; Indels 82; Gaps 12;
QY 11 QEELNOEKGQKEVLLRRCCQQLQEHGLAETRAEGLHQLEADHSRMRKREVS AHFHEVLR LK 70
Db 461 RDEQNEEK-----RR--QLQRLHEYET-----ELEDERNERALAAAKK--KLE 501
QY 71 DEMLSLSLHYSNALQEKELAAASRCRSLOEELYLKQELQ-----RAN 112
Db 502 GDLKDLQLEQADSAIKGREAIAKQLRKLAQMKDFQLEEDARASRDEIFATAKENEKKA 561
QY 113 MYSSCELELQEQSLRTASDQESGDEELNRLKEE-----NEKLSRSLTFSLAEKDI 161
Db 562 SLEADLMQLQEDLAAARAKQADLEKEELAEELASSLSGRNALQDEKRRLEARIAQ--- 618
QY 162 LEQSLDEARGSRQELVETHSLRERA-----VAAERQEQYWEKEQTLLQFQSKMKAC 215
Db 619 LEELEEEEGNMEAMSDRVKRTAQAEQLSNELATERSTAQNESARQO--LERQNKELRS 677
QY 216 QL-----YREKVNALQAVCELOKQDQAYSARDSAORETSQS-----LVEKD 258
Db 678 KLHEMGAVKSKFKSTIAALEAKIAQLEQVEQEAEREQAATKSLKQDKKLKILLQVE 737
QY 259 SLRRQVFELTDQ-----VCELRITQLRQLQAE 284
Db 738 DERKMAEQYKEQAEKGNARVKQLKRLQLEAEAE 770

Search completed: January 22, 2003, 08:57:35
Job time : 28.2218 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 11.8952 Seconds
(without alignments)
500.428 Million cell updates/sec

Title: US-09-767-215-2_COPY_126_420

Perfect score: 1462

Sequence: 1 ECLAGATGSLQELNQEKG.....TQLRQLQAEPPGVLKQEART 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1462	100.0	1004	10	US-09-767-215-2
2	1443	98.7	1138	10	US-09-767-215-5
3	369	25.2	1247	9	US-10-032-159A-8
4	364	24.9	319	9	US-10-032-159A-12
5	198	13.5	366	9	US-10-032-159A-2
6	198	13.5	366	9	US-10-032-159A-20
7	185	12.7	691	9	US-10-028-072-16
8	185	12.7	1597	9	US-10-017-216-6
9	185	12.7	1641	9	US-10-017-216-5
10	185	12.7	2055	9	US-10-017-216-4
11	184	12.6	868	9	US-09-884-001-19
12	183	12.5	1286	9	US-10-017-216-7
13	183	12.5	1958	12	US-10-028-946-4
14	183	12.5	2053	9	US-10-017-216-2
15	183	12.5	2054	12	US-10-028-946-2
16	177	12.1	2310	9	US-09-991-496-120
17	177	12.1	2310	10	US-09-874-923-120
18	174.5	11.9	2139	10	US-09-727-384-6
19	172	11.8	645	9	US-09-764-868-625

20	167.5	11.5	576	10	US-09-925-297-787	Sequence 787, App
21	166	11.4	689	9	US-10-108-605-305	Sequence 305, App
22	160.5	11.0	660	10	US-09-864-761-47959	Sequence 47959, A
23	160.5	11.0	1884	10	US-09-785-770A-17	Sequence 17, Appl
24	160.5	11.0	1907	10	US-09-785-770A-16	Sequence 16, Appl
25	159.5	10.9	909	10	US-09-925-299-988	Sequence 988, App
26	157.5	10.8	2125	10	US-09-919-172-29	Sequence 29, Appl
27	154	10.5	677	10	US-09-745-763-168	Sequence 168, App
28	153	10.5	303	10	US-09-925-297-740	Sequence 740, App
29	152	10.4	704	9	US-09-854-133-191	Sequence 191, App
30	152	10.4	704	10	US-09-738-973-191	Sequence 191, App
31	152	10.4	751	10	US-09-864-761-38419	Sequence 38419, A
32	151.5	10.4	374	10	US-09-925-302-711	Sequence 711, App
33	151.5	10.4	530	9	US-09-976-740-8	Sequence 8, Appli
34	151.5	10.4	530	10	US-09-962-055-8	Sequence 8, Appli
35	151.5	10.4	530	12	US-10-023-529-8	Sequence 8, Appli
36	151.5	10.4	530	12	US-10-023-523-8	Sequence 8, Appli
37	151.5	10.4	546	9	US-09-976-740-44	Sequence 44, Appl
38	151.5	10.4	546	12	US-10-023-529-44	Sequence 44, Appl
39	151.5	10.4	546	12	US-10-023-523-44	Sequence 44, Appl
40	151.5	10.4	557	9	US-09-976-740-5	Sequence 5, Appli
41	151.5	10.4	557	10	US-09-962-055-5	Sequence 5, Appli
42	151.5	10.4	557	12	US-10-023-529-5	Sequence 5, Appli
43	151.5	10.4	557	12	US-10-023-523-5	Sequence 5, Appli
44	151.5	10.4	1711	10	US-09-771-161A-219	Sequence 219, App
45	151.5	10.4	1711	10	US-09-771-161A-220	Sequence 220, App

ALIGNMENTS

RESULT 1
US-09-767-215-2
; Sequence 2, Application US/09767215
; Patent No. US20020081636A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-142001
; CURRENT APPLICATION NUMBER: US/09767,215
; PRIORITY FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/181,159
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-767-215-2

Query Match	100.0%;	Score 1462;	DB 10;	Length 1004;
Best Local Similarity	100.0%;	Pred. No. 2.6e-92;		
Matches 295;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ECLAGATGSLQELNQEKGKGVLLRRCCQLQEHGLGATRAEGLHOLEADHSRMKREVS	60	
Db	126	ECLAGATGSLQELNQEKGKGVLLRRCCQLQEHGLGATRAEGLHOLEADHSRMKREVS	185	
QY	61	AHFHEVLRLLKDEMLSLSHYSNALQEKELAAASRCRSLOEELYLKLOELQANWSSCELE	120	
Db	186	AHFHEVLRLLKDEMLSLSHYSNALQEKELAAASRCRSLOEELYLKLOELQANWSSCELE	245	
QY	121	LOEQSLRTASDQSGDEELNRLKEENEKLRSLTFSLAEKDILLEQSLDEARGSRQELVERI	180	
Db	246	LOEQSLRTASDQSGDEELNRLKEENEKLRSLTFSLAEKDILLEQSLDEARGSRQELVERI	305	
QY	181	HSILRRVAAERQROYWEKEQTLLQFOKSKMACOLYREKYNALQAOVCLEQKRDQAY	240	
Db	306	HSILRRVAAERQROYWEKEQTLLQFOKSKMACOLYREKYNALQAOVCLEQKRDQAY	365	
QY	241	SARSDAQREISQSLVEKDSLRQRQVFELTDQVCELRTQLRQLQAEPPGVULKQEART	295	


```

; PRIOR APPLICATION NUMBER: US 60/257,457
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 366
; TYPE: prt
; ORGANISM: Homo sapiens
US-10-032-159A-20

```

Query Match	13.5%	Score	198;	DB	9;	Length	366;
Best Local Similarity	26.7%	Pred.	No. 8.Be-07;				
Matches	67;	Conservative	48;	Mismatches	106;	Indels	30;
Gaps	6;						
QY	17	EKGQKEVLLRCCOOLQEHL	---	GLAETRAEGLHQLEADHS	-----	RMKREVSAAH	62
		: : : :		: : : :		:	
Db	112	ESGLTQLLMTTEVMKLOKKVQD	LTA	LLSKSDFKELRVKDSLLRKRKHQERVQRLKECEAG	171		
QY	63	FHEVRLRKDEWLSLSHYSNALQ	AKELAA	RCRSIQEELYLLKLOELORANMVSSCELELQ	122		
		: : :		: : :		: : :	
Db	172	SRELKCKKEENYDIAWRLAHQ	SEEEKGAALMRNRDLQLEIDOLKHS	LMAE--DCKVERK	229		
QY	123	EQ-SLRTASQESGEDELNR	KEE----	NEKLSRLTFSLAEK-----	DILEQSLDE	168	
		: :		: : : :		: : :	
Db	230	HTLKLRHAMEQRP	SOELLWELQEQKALQARVQ	ELASVQEGKLDRRSPY	IQVLEEDWRQ	289	
QY	169	ARGSRQELVERIHS	LRERAVAAERQ	EYBEEQETLLQFQKSMACOLY	REKYNALQAQ	228	
				: :		: : : :	
Db	290	ALRDHQEQANT	IFSLRKDLQGEARRLRCMEEK	EMFELQCLALRKDSKMYKDRI	FAILLQ	349	
QY	229	VCELOKERDQA	239				
		: :					
Db	350	MEEVAIERDOS	360				

```

RESULT 7
US-10-Q28-072-16
; Sequence 16, Application US/10028072
; Publication No. US2003000431A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIORITY APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIORITY APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIORITY APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIORITY APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIORITY APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIORITY APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17

```

```

1 FILE REFERENCE:
2 CURRENT APPLICATION NUMBER: US/10/028,072
3 CURRENT FILING DATE: 2001-12-19
4 PRIOR APPLICATION NUMBER: 60/049911
5 PRIOR FILING DATE: 1997-06-18
6 PRIOR APPLICATION NUMBER: 60/056974
7 PRIOR FILING DATE: 1997-08-26
8 PRIOR APPLICATION NUMBER: 60/059113
9 PRIOR FILING DATE: 1997-09-17
10 PRIOR APPLICATION NUMBER: 60/059115
11 PRIOR FILING DATE: 1997-09-17
12 PRIOR APPLICATION NUMBER: 60/059117
13 PRIOR FILING DATE: 1997-09-17
14 PRIOR APPLICATION NUMBER: 60/059122
15 PRIOR FILING DATE: 1997-09-17

```

[illegible]

; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090538
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07

Query Match 12.7%; Score 185; DB 9; Length 691;
Best Local Similarity 24.9%; Pred. No. 1.4e-05;
Matches 93; Conservative 55; Mismatches 129; Indels 96; Gaps 15;

QY 3 LAGATGSLQELN-----QKQKEVLLRRCQQL-----QHLGLAE----- 39
DB 149 LONQDESQQRNDMLQKLQLEGQVTELRSRVQELERATAREHTLMEQYKGISRS 208
QY 40 -----TRAEGHL-----QLEADHSRMKREYSAHFHEVLRKDBMLSLSHYSNA 83
DB 209 HGEITEERDILSRQGDHVARILEDDQITISEKVLTKEVELDRDLRTVYKALTREQEKL 268
QY 84 L-QEKELASRCRSLQEEYLLKQLQRANMYSCELELQ-----QSLRTASQSGDE- 137
DB 269 LGQLKEVQADKEQS-EAELQVAQENHNHL-----LDLKEAKSQEQAQRLKDKV 321
QY 138 -----ELNPLKEENKRLSLTSLAEK-DILQSLDEARGSRQELVERIH 181
DB 322 AQMKDTLGAQQRVAEPLKELQRLGAQELASASSQOKATLLGEELASAAAAARDRTIAELH 381
QY 182 SLRERAVAAE-----RQREYWEKEQTLLQFQKSMACOLYREKVNALQAOV 229
DB 382 --RSRLVAVNGRLAELGLHLKEKCQWSKERAGLLQ-----SVEAKDKILKLSAEI 433
QY 230 CELQK-----ERQVAYSARDSAQREISQSILVEKDSLRRQVVELTDVCELTQLRQAOBP 285
DB 434 LRLERAVQERTQNOVFKEELAREKDSLSVLQSEKRELTLSRLVQLQKEQLEQEK 493
QY 286 PGVL-----KQEAR 294
DB 494 QELLEYMRKLEAR 506

RESULT 8
US-10-017-216-6
; Sequence 6, Application US/10017216
; Patent No. US20020160483A1
; GENERAL INFORMATION:
; APPLICANT: KAPELLER-LIBERMANN, Rosana
; TITLE OF INVENTION: Kinase and Uses Therefor
; FILE REFERENCE: 10147-5701
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/242,429
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1597
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-017-216-6

Query Match 12.7%; Score 185; DB 9; Length 1597;
Best Local Similarity 26.5%; Pred. No. 3.6e-05;
Matches 88; Conservative 54; Mismatches 98; Indels 92; Gaps 16;

QY 11 QEELNQE-KGQKEVLLRRCQQLQEHGLAETRAEGL-HQLEADHSRM-----KREVS-A 61
DB 388 QEEMISELRQQKFYLETQAGKLEAQNRLKEOLEKISHQDHSDKSLLELETRLRVLSLE 447
QY 62 HPEVLRKDEMLSLSHYSNALQKE-----LAASRCRSLQEEYLLKQLQRANMVSS 116
DB 448 HEEQKLELRQLTELQ-----SLQERESQLTALQARA-ALESQLRQAATELEE-----TT 498
QY 117 CELELQEQSLRTASDQ-----ESGDELNRLKEENKRLSLTSLAEKDI 161
DB 499 AEAEIEIQALTAHRDEIQKFDALNRNSCTVITDLEBQLNQLTDEADNNAELNNQNFYLSKQ-- 556
QY 162 LEQSLDEARGSRQELVE---RIHSLRERAVAAERQREYWEKEQTLLQFQKSMACOLY 218
DB 557 ----LDEASGANDEIVQLRSEVDHLRREIT----EREMOLTSQKQTM---EALKTTCTTML 605
QY 219 REKVNALQA-----QVCELOKQERDOQVAYSARDSAQREI 250
DB 606 EEOVLDDLEALNDELLEKEREQWEAWSVLGDEKSQFECRVRELQRLMDLTEKQSRARADORI 665
QY 251 SOSLVEKDSLRRQVVELTDVCELTQLRQLQ 282
DB 666 TES-----RQVVELA--VKEHKAELALQ 687

RESULT 9
US-10-017-216-5
; Sequence 5, Application US/10017216
; Patent No. US20020160483A1
; GENERAL INFORMATION:
; APPLICANT: KAPELLER-LIBERMANN, Rosana
; TITLE OF INVENTION: 13245, A No. US20020160483A1 Human Myotonic Dystrophy Type P
; FILE REFERENCE: 10147-5701
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/242,429
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1641
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-017-216-5

Query Match 12.7%; Score 185; DB 9; Length 1641;
Best Local Similarity 26.5%; Pred. No. 3.8e-05;
Matches 88; Conservative 54; Mismatches 98; Indels 92; Gaps 16;

QY 11 QEELNQE-KGQKEVLLRRCQQLQEHGLAETRAEGL-HQLEADHSRM-----KREVS-A 61
DB 432 QEEMISELRQQKFYLETQAGKLEAQNRLKEOLEKISHQDHSDKSLLELETRLRVLSLE 491
QY 62 HPEVLRKDEMLSLSHYSNALQKE-----LAASRCRSLQEEYLLKQLQRANMVSS 116
DB 492 HEEQKLELRQLTELQ-----SLQERESQLTALQARA-ALESQLRQAATELEE-----TT 542
QY 117 CELELQEQSLRTASDQ-----ESGDELNRLKEENKRLSLTSLAEKDI 161
DB 543 AEAEIEIQALTAHRDEIQKFDALNRNSCTVITDLEBQLNQLTDEADNNAELNNQNFYLSKQ-- 600
QY 162 LEQSLDEARGSRQELVE---RIHSLRERAVAAERQREYWEKEQTLLQFQKSMACOLY 218
DB 601 ----LDEASGANDEIVQLRSEVDHLRREIT----EREMOLTSQKQTM---EALKTTCTTML 649
QY 219 REKVNALQA-----QVCELOKQERDOQVAYSARDSAQREI 250
DB 650 EEOVLDDLEALNDELLEKEREQWEAWSVLGDEKSQFECRVRELQRLMDLTEKQSRARADORI 709

```
QY 251 SOSLVKDSLRRQVFELTDVCELTQLRLQ 282
Db 710 TES-----RQVVELA--VKEHKAETLALQ 731

RESULT 10
US-10-017-216-4
; Sequence 4, Application US/10017216
; Patent No. US20020160483A1
; GENERAL INFORMATION:
; APPLICANT: KAPPELLER-LIBERMANN, Rosana
; TITLE OF INVENTION: 13245, A No. US20020160483A1el Human Myotonic Dystrophy Type Prob
; FILE OF INVENTION: Kinase and Uses Therefor
; CURRENT APPLICATION NUMBER: US/10/017,216
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/242,429
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2055
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-017-216-4

Query Match 12.7%; Score 185; DB 9; Length 2055;
Best Local Similarity 26.5%; Pred. No. 4.8e-05;
Matches 88; Conservative 54; Mismatches 98; Indels 92; Gaps 16;

QY 11 QEELNOE-KGQEVLLRRQVQLQGLAEAEGL-HOLEADHSM-----KREVS-A 61
Db 846 QEEMISELRQKQFYLETQAGKLEAQNRLKEEQLKISHQDSDKSLLELTFLREVSLE 905
QY 62 HFHEVLRLKDEMLSLSHYSNALQKE-----LAASRCRSLQEBELVLLKQELQANMVS 116
Db 906 HEQKLELKRQLTELQ-----SLQERESOLTALQARA-ALESOLRQAKTELEE---TT 956
QY 117 CELELQESQSLRTASDQ-----ESGDELNRLKEENKLSLTFSLAEKDI 161
Db 957 AEAEETEQALTAHRDETQKFDALNRNCTVTITDLEQLNQLTQEDNAELNNQNFYLSKQ-- 1014
QY 162 LEQSLDEARSGROELVE---RIHSLRERAVAAERQEQVWEKEQTILLOFQKSKMACOLY 218
Db 1015 -----LDASGANDIEVQLRSEVDHLREIT-----EREMQLTSQKQIM---EALKTTCTML 1063
QY 219 REKYNALQA-----QVCELOKEDQAYSARDSAQREI 250
Db 1064 EEQVLDLEALNDELLEKQERQWEAWRSVLGDEKSFQFCRVRELQRLMLDTEKQSRARADQRI 1123
QY 251 SOSLVKDSLRRQVFELTDVCELTQLRLQ 282
Db 1124 TES-----RQVVELA--VKEHKAETLALQ 1145

RESULT 11
US-09-884-001-19
; Sequence 19, Application US/09884001
; Publication No. US20020182656A1
; GENERAL INFORMATION:
; APPLICANT: Bird, Timothy A.
; APPLICANT: Peschon, Jacques J.
; APPLICANT: Sims, John E.
; APPLICANT: Virca, G. Duke
; APPLICANT: Willis, Cynthia R.
; TITLE OF INVENTION: Methods for Regulating Vascularization Using GEF
; FILE OF INVENTION: Containing NEK-Like Kinase (GNK)
; FILE REFERENCE: Immurex GNK/sgNK PCT
; CURRENT APPLICATION NUMBER: US/09/884,001
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/113,003
; PRIOR FILING DATE: 1998-12-18
```

```
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 868
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-884-001-19

Query Match 12.6%; Score 184; DB 9; Length 868;
Best Local Similarity 25.9%; Pred. No. 2.1e-05;
Matches 89; Conservative 52; Mismatches 117; Indels 86; Gaps 14;

QY 12 BELNOE-KGQEVLLRRQVQLQGLAEAEGL-HOLEADHSMKREVSAAHFHEVL 67
Db 51 EQLHQAERQEVLLARAVQEKALVREKAALEVR---LQAVDRDQLAEQLQ-----L 102
QY 68 RLKDEMLSLSL-----HYSNALQKEKELAAASRCRSLQEBELVLLKQEL 108
Db 103 SSAKELLESSLFEAQOQNSVIEVTGKQLEVIQTVTQAKVIGQEVRCLELDTSTERSQA 162
QY 109 ORANMVSCELEQESQSLRTASDQESG--DEELNRLKEENKLSLTFSLAEKDILEQSL 166
Db 163 EQRDAAARQLAQAEQCKTALQOQKAHEKEVNLKWEKERSW-----HQELAKAL 217
QY 167 DEARSGROELVERIHSRLERAVAAERQEQVWEKEQTILLOFQKSKMACOLY 208
Db 218 ESLEREKMELEMLRLKEQQTQEMEAIAQAEERERTQAESALCOMOLETEKERVSLLETLIQT 277
QY 209 QK--SKMACOLYR-----EKVNALQAVCELOKE-RDOAYSARD---SAQRE 249
Db 278 QKELADASQQLERLQDMKVKOKLEQETTGILQTLQQAQRELKEMARQHRDDLALQEE 337
QY 250 SOSLVKDSLRRQVFELTDVCELTQLRLQ 282
Db 338 SSSLQDRKMDLQKQVEDLKSQVADDSQRLVQEVQEKLERETQ 381

RESULT 12
US-10-017-216-7
; Sequence 7, Application US/10017216
; Patent No. US20020160483A1
; GENERAL INFORMATION:
; APPLICANT: KAPPELLER-LIBERMANN, Rosana
; TITLE OF INVENTION: 13245, A No. US20020160483A1el Human Myotonic Dystrophy Type P
; FILE OF INVENTION: Kinase and Uses Therefor
; FILE REFERENCE: 10147-5701
; CURRENT APPLICATION NUMBER: US/10/017,216
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/242,429
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1286
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-216-7

Query Match 12.5%; Score 183; DB 9; Length 1286;
Best Local Similarity 26.2%; Pred. No. 3.9e-05;
Matches 87; Conservative 55; Mismatches 98; Indels 92; Gaps 16;

QY 11 QEELNOE-KGQEVLLRRQVQLQGLAEAEGL-HOLEADHSM-----KREVS-A 61
Db 79 QEEMISELRQKQFYLETQAGKLEAQNRLKEEQLKISHQDSDKSLLELTFLREVSLE 138
QY 62 HFHEVLRLKDEMLSLSHYSNALQKE-----LAASRCRSLQEBELVLLKQELQANMVS 116
Db 139 HEQKLELKRQLTELQ-----SLQERESOLTALQARA-ALESOLRQAKTELEE---TT 189
QY 117 CELELQESQSLRTASDQ-----ESGDELNRLKEENKLSLTFSLAEKDI 161
Db 190 AEAEETEQALTAHRDETQKFDALNRNCTVTITDLEQLNQLTQEDNAELNNQNFYLSKQ-- 247
```



```
Db 907 HEEQKLELKRQTELOL-----SLQERESOLTALQAARA-ALBSQLRQAkteLEE-----TT 957
QY 117 CELELOEQSLRTASDQ-----ESGDEELNRLKEENEKLSLTFSLAEKDI 161
Db 958 AEAEIEIQALTARDEIQRKFDALRNSCTVITDLEEQNLQLTEDNAELNNQNFYLSKO-- 1015
QY 162 LEQSLDEARGSRQELVE---RIHSLRERAVAAERQEQYWEKEQTLLQFQKSKMACQLY 218
Db 1016 -----LDEASGANDEIVQLRSEVDHLRREIT-----EREMOLTSQKQTM---EALKTTCTML 1064
QY 219 REKVNALQA-----QVCELOKERDQAYSARDSAQREI 250
Db 1065 EQQVMDLEALNDELLEKQWEAWRSVLGDEKSQFECRVRELQRLMDTEKQSRARADORI 1124
QY 251 SOSLVEKDSLRQVFELTDQVCELRTQLRLO 282
Db 1125 TES-----ROVVELA---VREHKAIEILALQ 1146
```

Search completed: January 22, 2003, 08:52:51
Job time : 17.8952 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 22.6008 seconds
(without alignments)
1254.807 Million cell updates/sec

Title: US-09-767-215-2_COPY_126_420

Perfect score: 1462
Sequence: 1 ECLAGAGSLQELNQKQ.....TQLRQLQAEPPGVKQEQART 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	202.5	13.9	1407	1	S28589
2	194	13.3	2101	2	A42184
3	192	13.1	2168	2	T30171
4	191.5	13.1	1690	2	T13030
5	188.5	12.9	2663	1	S28261
6	188	12.9	1133	2	T22976
7	187.5	12.8	1256	2	T08621
8	186.5	12.8	2442	2	G02520
9	185.5	12.7	4574	2	A59404
10	185.5	12.7	4684	2	S58420
11	185	12.7	1597	2	UC5837
12	185	12.7	3187	2	S21801
13	184.5	12.6	1999	1	S21801
14	184	12.6	746	2	T47237
15	182.5	12.5	1286	2	T16507
16	181.5	12.4	3225	2	T52300
17	181.5	12.4	4687	1	A39638
18	181	12.4	1976	2	A59252
19	181	12.4	2954	2	T14156
20	180.5	12.3	1927	2	A59236
21	180.5	12.3	1959	1	A33977
22	179.5	12.3	3259	1	A56539
23	178.5	12.2	1218	2	T14265
24	178.5	12.2	1940	1	S04095
25	177	12.1	583	2	C84788
26	177	12.1	1992	2	A47297
27	175.5	12.0	1738	2	T14867
28	175.5	12.0	1937	2	T18055
29	175	12.0	1964	2	A59282

30 175 12.0 2007 1 B43402
31 174.5 11.9 638 2 T03791
32 173.5 11.9 1475 2 T33318
33 173.5 11.9 1549 1 A40691
34 173.5 11.9 1940 1 A24922
35 173.5 11.9 2139 2 T18296
36 173 11.8 924 2 S06117
37 172.5 11.8 899 2 I38153
38 172.5 11.8 1732 2 T14039
39 172.5 11.8 1790 2 S67593
40 172.5 11.8 2020 2 T21174
41 172 11.8 1961 1 A61231
42 171.5 11.7 845 2 I48176
43 171.5 11.7 946 2 S28061
44 171.5 11.7 993 2 S49461
45 171.5 11.7 1017 2 PC4035

ALIGNMENTS

RESULT 1
S28589

trichohyalin - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999

C:Accession: S28589

R:Fietz, M.J.; Rogers, G.E.

submitted to the EMBL Data Library, December 1992

A:Description: Examination of the gene encoding rabbit trichohyalin.

A:Reference number: S28589

A:Accession: S28589

A:Molecule type: DNA

A:Residues: 1-1407 <FIE>

A:Cross-references: EMBL:Z19092; NID:gl746; PIDN:CAA79519.1; PID:gl747

C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root
Covalent modifications to this protein include conversion of arginine to citrulline a

C:Genetics:

A:Introns: 46/3

C:Superfamily: trichohyalin; calmodulin repeat homology

C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat

F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 13.9%; Score 202.5; DB 1; Length 1407;
Best Local Similarity 27.7%; Pred. No. 0.0011;
Matches 97; Conservative 48; Mismatches 108; Indels 97; Gaps 13;

Qy 11 QEELNQEKQK---EVLLRQCQLQELGLAETRAEGLHQL-----EADHSRMKRE 58

Db 641 EQELRQERERKLREEEQLLRREQ-----ELRQERERKLREEEQLLRQERERLRQ 692

Qy 59 VSAHFVRLKDEMLSLSHYSNALQEKELAAASCRSLQEBLYLLKQELQANMVSCE 118

Db 693 ERA-----RKLREEQLL-----RQEEQLRQERERKLREEEQLLRQERDRK 741

Qy 119 LELOQSILRTASDQSGDEELNRLKEENKRLSLTFLAEDILFQSLDEARGSRQE--- 175

Db 742 LREEQLL-----QESEERLRQERQELRRDRKFRREEQLLRQERERLRQERER 796

Qy 176 -----LVERIHSLR-----ERAAERQR-----EQYWEK 201

Db 797 KLREEEQLLRQERERLRQERERKLREEEQLLRQERERLRQERERKLREEEQLLRQE 856

Qy 202 EOTLLQFQKSNACOLYREKYNALQAVCELOKRD-----QAYSARSA 246

Db 857 EQLRQERARKL-----REEEQLLRQERQERDRKLREEEQLLRQERQERDRK 911

Qy 247 QRETSQSL--VEKDSLRQVFE--LTDQVCELRQLRQLOAEPPGVKQEQ 292

Db 912 LREEEQLLRQERERLRQERERKLREEEQLLRQERERLRQERERARKLREE 961

RESULT 2

myosin heavy chain
outer dense fiber
hypothetical prote
trichohyalin - she
myosin heavy chain
myosin heavy chain
myosin heavy chain
gene retif protein
protein kinase (EC
transport protein
hypothetical prote
myosin heavy chain
synaptonemal compl
SCPI protein - rat
synaptonemal compl
cell-cycle-depende

Db 2581 EARRRQHEAEGVRRKQEELEQLEQQ 2606

RESULT 10

A59404

pectin [imported] - human

C:Species: Homo sapiens (man)

C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 19-Apr-2002

C:Accession: C59404; A59404

R:Liu, C.G.; Maercker, C.; Castanon, M.J.; Hauptmann, R.; Wiche, G.

Proc. Natl. Acad. Sci. U.S.A. 93, 4278-4283, 1996

A:Title: Human plectin: organization of the gene, sequence analysis, and chromosome localization

A:Reference number: C59404; MUID:96210632; PMID:8633055

A:Accession: C59404

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-4684 <STO>

A:Cross-references: GB:CAA911196; NID:g1296662; PIDN:CAA91196.1

C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S

Query Match 12.7%; Score 185.5; DB 2; Length 4684;

Best Local Similarity 27.0%; Pred. No. 0.032;

Matches 88; Conservative 41; Mismatches 134; Indels 63; Gaps 11;

QY 10 LOELNQEKQKEVLLRRCCQQLQEHGLGLAETRAEGLHLQLEADHSRMRKREVSAMHFEVLR 69

Db 2403 LOBEAEKMQVAEEAARLVAAQEAARLRQLAEEDLAQORALAEKMLKEMQAVQEAETRL 2462

QY 70 KDEMLSLHYSNALQEKELASRCSLSIOBELYLLKQEL--QPNAMVSSCELELQEQ--- 124

Db 2463 KAEAEQLQ-----QOEKLAQEQARRLEQEKQMAQQLAAEETQGFQRTLEAERQRLQEM 2515

QY 125 -----SLRTAS---DOESGDEELNRLKEE---NEKLSRLTFSLAEKDILQESLDEAR 170

Db 2516 SAFAERLKLVAEMSAQARAEDAQRFKQAEIEGKELHRTLETATQEKVTLVOTLEIOR 2575

QY 171 GSRQELVRIHSRRAVAERQREYWEKEQTLLQFQKSMAC-----QLYREKYNALQ 226

Db 2576 QQSDHDAER---LREAIAELEREKEKL--QOEAKLLQLKSEEMOTVQEQQLQETQALQ 2630

QY 227 AQVCE-----LQER-----DOAYSARSAQREISQSILVEKDSL----- 260

Db 2631 SFLSEKDSLQLRERFTEQEKAKLEQLFQDEKVAQAQLREEQQRQOQOQEQRLVASME 2690

QY 261 --RRQVFELTDQVCELTOLRLOAE 284

Db 2691 EARRRQHEAEGVRRKQEELEQLEQQ 2716

RESULT 11

S68420

citron - mouse

C:Species: Mus musculus (house mouse)

C>Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999

C:Accession: S68420

R:Madaule, P.; Furuyashiki, T.; Reid, T.; Ishizaki, T.; Watanabe, G.; Morii, N.; Narumiy

FEB8 Lett. 377, 243-248, 1995

A:Title: A novel partner for the GTP-bound forms of rho and rac.

A:Reference number: S68420; MUID:96128238; PMID:8543060

A:Accession: S68420

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1597 <MAD>

A:Cross-references: EMBL:U39904; NID:g1079733; PIDN:AAC52341.1; PID:g1079734

C:Superfamily: protein kinase C zinc-binding repeat homology; plectstrin repeat homology

C:Keywords: alternative splicing

F:931-979/Domain: protein kinase C zinc-binding repeat homology <K22>

Query Match 12.7%; Score 185; DB 2; Length 1597;

Best Local Similarity 26.5%; Pred. No. 0.011;

Matches 88; Conservative 54; Mismatches 98; Indels 92; Gaps 16;

QY 11 QEELNQE-KGQKEVLLRRCCQQLQEHGLGLAETRAEGL-HOLEADHSRM-----KREVS-A 61

Db 388 QEEMISELRQQRFFYLETQAGKLEAQRNKLKEEQEIKSHODHSDKSLLELETRLEVSLE 447

QY 62 HFHEVLRLKDEMLSLHYSNALQEK-----LAASRCRSLQEEYLLKQELQRANMVSS 116

Db 448 HEEQKLELKRQLTELQ-----SLQRESOLTALQARA-ALSQLRQAQKTELEE-----TT 498

QY 117 CELELQEQSLRTASDQ-----ESGDEELNRLKEENEKLSLTFSLAEKDI 161

Db 499 AEAEIEIQALTAHRDEIQRKFDALRNSCIVITDLEQLNQLTEDNAELNNQNFYLSKQ-- 556

QY 162 LQOSLDARGSRQELVE---RTHSLRRAVAARQREYWEKEQTLLQFQSKMACQLY 218

Db 557 ----LDEASGANDIEVQLRSEVDHLRREIT-----EREMQLTSQKQTM---EALKTTCTML 605

QY 219 REKVNALQA-----OVCELOKQERDQAYSARDSAQREI 250

Db 606 EEOVLDEALNDELLEKEKQWAMRSLVGDQSKQFCFVRLEIQLRMLDTEKQSRARADORI 665

QY 251 SQSLVEKDSLRRQVFELTDQVCELTQLRLQ 282

Db 666 TES-----RQVVELA--VKEHKAIEILALQ 687

RESULT 12

JC5837

364K Golgi complex-associated protein - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000

C:Accession: JC5837

R:Toki, C.; Fujiwara, T.; Sohma, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.

Cell Struct. Funct. 22, 565-577, 1997

A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein

A:Reference number: JC5837; MUID:98093490; PMID:9431462

A:Accession: JC5837

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-3187 <TOK>

A:Cross-references: DBJ:D25543; NID:g516825; PIDN:BAAO5026.1; PID:g516826

C:Comment: This protein plays a role in the formation and maintenance of the character

C:Superfamily: giantin

F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status pred

F:3165-3187/Domain: membrane anchor #status predicted <MAD>

Query Match 12.7%; Score 185; DB 2; Length 3187;

Best Local Similarity 25.3%; Pred. No. 0.023;

Matches 82; Conservative 54; Mismatches 98; Indels 90; Gaps 13;

QY 7 IGSLOEELNQERQ-----KEVLLRRCCQ-----QLQEHGLGLAETRAEGLHQ 47

Db 200 IGALQTLQSLQTAQEAQAQKLRYVMQRKLEEBHEEALLGRAQVVDLLQKELTSAEQRNQDLSQ 259

QY 48 ----LEADHSRMRKREVSAMHFEVLRKDEMLSLHYSNALQEK-----ELAASRCRSLQEE 100

Db 260 QLQLLEAEHSTLRNTMEARQE-----SKILMEKVELEMAERK-----EE 299

QY 101 LVLLKQELQRANMVSSCELEEQSLRTASDQ-----SGDEEL-----NRLKE 144

Db 300 LYQLQQLERAGQ-AQAELEMQVGLTQQRHETEMEKTACISLLQKNEQELASACDALKE 358

QY 145 ENKLSRLTFSLAEK--DILEQSLDEARGSRQELVRIH--SLRRAVAARQREYWE 200

Db 359 ENSKLLQEQEQEAQSAQAQLQLEDELEQKSKSEISQFVNKPNLEKHETSSQTLSDPVYNE 418

QY 201 KGQTLQFQKSMACQLYREKYNALQAVCELOKQERDQAYSARDSAQREISQSLVEKDSL 260

Db 419 GVQAVNE-----ESVASLQKRVLELENEKG-----ALLLSLELEEL 455

QY 261 RRQVFELTDQVCELTQLRLOAE 284

Db 456 RAENEKLCRSRITLLLEAQNAGEAD 479

Db 162 EKLSAKANDIFWVTKDLHDKNEELTSFRMEYVTKLSE---ANREKKALEEKLEKYKNDM 218
QY 109 QRANMVSCELELQEOSLRTAS-----DOESGDEELNRLKEENKLRSL-TFSLAEKDI 161
Db 219 KENDRKS---LELNKEQVTTQNVLSEVRQLSAHFELTPVRKNASKIRELDEYHQLSAKV 275
QY 162 LEOSLDEAR-----GSRQELV-----ERHSLRERAVAA---ERQREQYWEEKQT 204
Db 276 IEESMNDLKIKNETLTKELSDKTELVKMKNEELEDLRQTTTASLGDSQATKYLHEENMK 335
QY 205 LLOFQKSKMACOLY--REKVNALQAVCELOKQERDOAYSARDSAQREISQSLV-EKDSLR 261
Db 336 LTR-QKADIRCELEARRKVEGFDKLQBLEKERDDAL-ADVOKIREVKRNVERELQSLT 393
QY 262 ROVFELTDQVCELRLOT 278
Db 394 SLMAERDEQIEELKTKM 410

Search completed: January 22, 2003, 08:54:36
Job time : 31.6008 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 11.4987 Seconds
(without alignments)
1064.082 Million cell updates/sec

Title: US-09-767-215-2_COPY_126_420

Perfect score: 1462
Sequence: 1 ECLAGAGLSQELNQKQGQ.....TQLRLQAEPPVLKQEART 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1462	100.0	1004	1 CARE_HUMAN	Q9bx16 homo sapien
2	1166	79.8	999	1 CARE_MOUSE	Q99kf0 mus musculus
3	396.5	27.1	1147	1 CARE_HUMAN	Q9bx17 homo sapien
4	329.5	22.5	1032	1 CARE_HUMAN	Q9bwt7 homo sapien
5	322.5	22.1	1021	1 CARE_MOUSE	P58660 mus musculus
6	283.5	19.4	536	1 CAR9_RAT	Q9ep90 rattus norv
7	274.5	18.8	536	1 CAR9_HUMAN	Q9h257 homo sapien
8	202.5	13.9	1407	1 TRH9_RABIT	P37709 oryctolagus
9	188.5	12.9	2663	1 CENE_HUMAN	Q02224 homo sapien
10	186.5	12.8	1976	1 MYH4_RAT	Q9jlt0 rattus norv
11	186	12.7	4473	1 PLE1_CRIGR	Q9j155 cricetus
12	185.5	12.7	4684	1 PLE1_HUMAN	Q15149 homo sapien
13	185	12.7	1597	1 CTRO_MOUSE	P49025 mus musculus
14	183	12.5	1286	1 CTRO_HUMAN	O14578 homo sapien
15	182	12.4	1976	1 MYH4_BOVIN	Q27991 bos taurus
16	181.5	12.4	4687	1 PLE1_RAT	P30427 rattus norv
17	181	12.4	1976	1 MYH4_HUMAN	P35580 homo sapien
18	180.5	12.3	1959	1 MYH9_CHICK	P14105 gallus gall
19	179.5	12.3	1938	1 MYH8_HUMAN	Q9ukx3 homo sapien
20	178.5	12.2	1940	1 MYH3_HUMAN	P11055 homo sapien
21	177	12.1	1961	1 MYH9_RAT	Q62812 rattus norv
22	176.5	12.1	976	1 SCPI_HUMAN	Q15431 homo sapien
23	176	12.0	2230	1 GOG4_HUMAN	Q13439 homo sapien
24	175.5	12.0	1937	1 MYH8_HUMAN	P13535 homo sapien
25	174	11.9	1939	1 MYH4_HUMAN	Q9V623 homo sapien
26	174	11.9	1960	1 MYH9_HUMAN	P35579 homo sapien
27	173.5	11.9	1349	1 TRH9_SHEEP	P22793 ovis aries
28	173.5	11.9	1940	1 MYH3_RAT	P12847 rattus norv
29	173.5	11.9	3210	1 CENE_HUMAN	P49454 homo sapien
30	172.5	11.8	1790	1 US01 YEAST	P25386 saccharomyc
31	171.5	11.7	845	1 SCPI_MESAU	Q60563 mesocricetu
32	171.5	11.7	993	1 SCPI_MOUSE	Q02409 mus musculus
33	171.5	11.7	997	1 SCPI_RAT	Q03410 rattus norv

RESULT 1
CARE_HUMAN
ID CARE_HUMAN STANDARD; PRT; 1004 AA.
AC Q9BX16; Q9BVBS;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 14 (CARD-containing MAGUK protein
DE 2) (Carma 2).
GN CARD14 OR CARMA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21192234; PubMed=11278692;
RA Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA Srinivasula S.M., Merriam S., DiStefano P.S., Alnemri E.S.;
RT "CARD11 and CARD14 are novel caspase recruitment domain
RT (CARD)/membrane-associated guanylate kinase (MAGUK) family members
RT that interact with Bcl10 and activate NF-kappaB.";
RL J. Biol. Chem. 276:11877-11882(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21255663; PubMed=11356195;
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RT "Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10
RT phosphorylation and NF-kappaB activation.";
RL FEBS Lett. 496:121-127(2001).
RN [3]
RP ERRATUM.
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RL FEBS Lett. 505:198-198(2001).
RN [4]
RP SEQUENCE OF 1-740 FROM N.A.
RC TISSUE=Cervix, and Colon;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the
CC phosphorylation of Bcl10.
CC -!- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD
CC interaction.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Expressed in placenta. Also detected in HeLa
CC S3 cells, but not in the other cancer cell lines tested.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- CAUTION: Supposed to contain a SH3 domain which is not detected by
CC PROSITE, Pfam or SMART.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF322642; AAG53403.1; -;
DR EMBL; AY032927; AAK54453.1; -;
DR EMBL; BC018142; AAH18142.1; -;
DR EMBL; BC001326; AAH01326.1; ALT INIT.
DR InterPro: IPR000619; Guanylate_kin.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50209; CARD; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Coiled coil.
FT DOMAIN 15 107 CARD.
FT DOMAIN 128 409 COILED COIL (POTENTIAL).
FT DOMAIN 568 658 PDZ.
FT DOMAIN 858 990 GUANYLATE KINASE.
FT CONFLICT 619 671 DYEASEPLEKAVLEDTLLEAVGLLRVDGFCCLSKVNVTD
FT GYKRLQDLLEAK -> SRAPLLSPGLMGTVAAAGVQAD
FT FTSPRRCRSTLGSALSADVKS AHL (IN REF. 2;
FT AAH01326).
SQ SEQUENCE 1004 AA; 113299 MW; 7469B8B56BE06073 CRC64;

Query Match 100.0%; Score 1462; DB 1; Length 1004;
Best Local Similarity 100.0%; Pred. No. 3.8e-71;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ECLAGAGISLQELNQEKGQKQEVLLRRCQQLQEHGLAETRAEGLHQLADHSMKREVS 60
Db 126 ECLAGAGISLQELNQEKGQKQEVLLRRCQQLQEHGLAETRAEGLHQLADHSMKREVS 185
QY 61 AHFHEVLRLKDEMLSLSHYSNALQEKELAAASRCRSQLEELYLLKQELQORANMVSSCELE 120
Db 186 AHFHEVLRLKDEMLSLSHYSNALQEKELAAASRCRSQLEELYLLKQELQORANMVSSCELE 245
QY 121 LOEQSLRTASDOESGDGEELNRLKEENKRLSLTFSLEAKDILEQSLDEARGSRQELVERI 180
Db 246 LOEQSLRTASDOESGDGEELNRLKEENKRLSLTFSLEAKDILEQSLDEARGSRQELVERI 305
QY 181 HSLRERAAERQEQWEKEQTLQFQSKMACQLYREKYNALQAOVCLOKQERDQAY 240
Db 306 HSLRERAAERQEQWEKEQTLQFQSKMACQLYREKYNALQAOVCLOKQERDQAY 365
QY 241 SARDAQREISQSLVEKDSLRQVFEITDQVCELTQRLQOLQAEPPGVVLKQEA 295
Db 366 SARDAQREISQSLVEKDSLRQVFEITDQVCELTQRLQOLQAEPPGVVLKQEA 420

RESULT 2
CARE_MOUSE STANDARD; PRT; 999 AA.
AC Q99KF0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 14 (Bcl10-interacting MAGUK protein
DE 2) (Bimp2).
GN CARD14 OR BIMP2.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=21391892; PubMed=11387339;
RA McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A.,
RA Li Q., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W.,
RA Nunez G.;
RA "Bimp1, a MAGUK family member linking protein kinase C activation to

RT Bcl10-mediated NF-kappa B induction.";
RL J. Biol. Chem. 276:30589-30597(2001).
RN [2]
RP SEQUENCE OF 82-743 FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC !- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD
CC interaction (By similarity).
CC !- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC !- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC !- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC !- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC !- CAUTION: Supposed to contain a SH3 domain which is not detected by
CC PROSITE, Pfam or SMART.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF363457; AAK60137.1; -;
DR EMBL; BC004692; AAH04692.1; -;
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR PROSITE; PS50209; CARD; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Coiled coil.
FT DOMAIN 15 107 CARD.
FT DOMAIN 125 411 COILED COIL (POTENTIAL).
FT DOMAIN 572 655 PDZ.
FT DOMAIN 854 986 GUANYLATE KINASE.
FT CONFLICT 736 743 QAAQQLLA -> HLEEDHRS (IN REF. 2).
SQ SEQUENCE 999 AA; 113496 MW; D18350DA12430255 CRC64;
Query Match 79.8%; Score 1166; DB 1; Length 999;
Best Local Similarity 79.9%; Pred. No. 2.1e-55;
Matches 234; Conservative 29; Mismatches 30; Indels 0; Gaps 0;
QY 1 ECLAGAGISLQELNQEKGQKQEVLLRRCQQLQEHGLAETRAEGLHQLADHSMKREVS 60
Db 126 ECLAGAGISLQELNQEKGQKQEVLLRRCQQLQEHGLAETRAEGLHQLADHSMKREVS 185
QY 61 AHFHEVLRLKDEMLSLSHYSNALQEKELAAASRCRSQLEELYLLKQELQORANMVSSCELE 120
Db 186 AHFHEVLRLKDEMLSLSHYSNALQEKELAAASRCRSQLEELYLLKQELQORANMVSSCELE 245
QY 121 LOEQSLRTASDOESGDGEELNRLKEENKRLSLTFSLEAKDILEQSLDEARGSRQELVERI 180
Db 246 LOEQSLRTASDOESGDGEELNRLKEENKRLSLTFSLEAKDILEQSLDEARGSRQELVERI 305
QY 181 HSLRERAAERQEQWEKEQTLQFQSKMACQLYREKYNALQAOVCLOKQERDQAY 240
Db 306 HSLRERAAERQEQWEKEQTLQFQSKMACQLYREKYNALQAOVCLOKQERDQAY 365
QY 241 SARDAQREISQSLVEKDSLRQVFEITDQVCELTQRLQOLQAEPPGVVLKQEA 293
Db 366 SARDAQREISQSLVEKDSLRQVFEITDQVCELTQRLQOLQAEPPGVVLKQEA 418
RESULT 3
CARE_HUMAN STANDARD; PRT; 1147 AA.
ID CARB_HUMAN
AC Q9BXL7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

15-JUN-2002 (Rel. 41, Last annotation update)
 Caspase recruitment domain protein 11 (CARD-containing MAGUK protein
 3) (Carna 1).
 CARD11 OR CARMAL.
 Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=21192234; PubMed=1127862;
 Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L., Poyet J.-L.,
 Srinivasula S.M., Merriam S., Distefano P.S., Alnemri E.S.;
 "CARD11 and CARD14 are novel caspase recruitment domain
 (CARD)/membrane-associated guanylate kinase (MAGUK) family members
 that interact with Bcl10 and activate NF-kappaB.";
 J. Biol. Chem. 276:11877-11882(2001).
 [2]
 SEQUENCE FROM N.A.
 MEDLINE=21255663; PubMed=11356195;
 Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschoopp J.;
 "Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10
 phosphorylation and NF-kappaB activation.";
 FEBS Lett. 496:121-127(2001).
 [3]
 ERRATUM.
 Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschoopp J.;
 FEBS Lett. 505:198-198(2001).
 -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the
 phosphorylation of Bcl10.
 -!- SUBUNIT: CARD11 and Bcl10 bind to each other by CARD-CARD
 interaction.
 -!- SUBCELLULAR LOCATION: Cytoplasmic.
 -!- TISSUE SPECIFICITY: Detected in adult peripheral blood leukocytes,
 thymus, spleen and liver. Also found in promyelocytic leukemia HL-
 60 cells, chronic myelogenous leukemia K562 cells, Burkitt's
 lymphoma Raji cells and colorectal adenocarcinoma SW480 cells. Not
 detected in HeLa S3, Molt-4, A549 and G431 cells.
 -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
 -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
 -!- CAUTION: Supposed to contain a SH3 domain which is not detected by
 PROSITE, Pfam or SMART.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AF322641; AAG53402.1; --
 Genew; HGNC:16393; CARD11.
 InterPro; IPR000619; Guanylate_kin.
 InterPro; IPR001478; PDZ.
 SMART; SM00228; PDZ; 1.
 PROSITE; PS02029; CARD; 1.
 PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
 PROSITE; PS00502; GUANYLATE_KINASE_2; FALSE_NEG.
 PROSITE; PS0106; PDZ; FALSE_NEG.
 Coiled coil.
 KW DOMAIN 11 103 CARD.
 FT DOMAIN 123 442 COILED COIL (POTENTIAL).
 FT DOMAIN 673 748 PDZ.
 FT DOMAIN 966 1133 GUANYLATE KINASE.
 FT CONFLICT 808 808 P -> L (IN REF. 2).
 SQ SEQUENCE 1147 AA; 913A4B01502B36CC CRC64;
 Query Match 27.1%; Score 396.5; DB 1; Length 1147;
 Best Local Similarity 34.0%; Pred. No. 2.2e-14;
 Matches 102; Conservative 66; Mismatches 109; Indels 23; Gaps 6;

Oy 3 LAGATGSGEELNQKQKEVLLRRCCQLQ-EHLGLAETRAEGLHQLADHSRMKREYSA 61
 Db 124 LMNEVIKQQQAKKALQRCCELLARLQLEDEKQMTLTRVE-LLTFTQERYKMKERDS 182
 Oy 62 HFHEVRLRLKQDMLSLSHYSNALQEKELAAASRCRSLQELVLLKQELORANVSSCELEL 121
 Db 183 YNDELVKYKDDNYNIAMRYAQLESEKNAVMKRSRDLQLEIDQLKHLNKK--MEECKLE- 239
 Oy 122 QEQSLRTASDQES--GDEELNRLKEENKRLSLTSL-----AEKDILE 163
 Db 240 RNQSLKLNKNDIENRPKKEQVLELERENEMLTKNQELQSIQAGKRSLPDSKAILDILE 299
 Oy 164 QSLDEARGSRQELVERIHSIRERAVAAERQRYWEKEQTLLQFKQKMAQOLYREKVN 223
 Db 300 HDRKALEDRQELVNRINYNLQEAQAEELRDYKLEEKDELKCTLGKDCMKYHRMN 359
 Oy 224 ALQAOVCLEKQERDAYSAQDSQAEISQSLVEKDSLQVPELTDQVCELTQRLQLOA 283
 Db 360 TVMLQLEEVERDQAFHSRDEAQTQYSQCLIEKDKYRKQRELEKNDKEMRIEMVRREA 419
 RESULT 4
 CARA_HUMAN
 ID CARA_HUMAN STANDARD; PRT: 1032 AA.
 AC Q9BWT7; Q9UGR5; Q9UGR6; Q9Y3H0;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Caspase recruitment domain protein 10 (CARD-containing MAGUK protein
 3) (Carna 3).
 GN CARD10 OR CARMA3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2129987; PubMed=11259443;
 RA Wang L., Guo Y., Huang W.-J., Ke X., Poyet J.-L., Manji G.A.,
 Merriam S., Glucksmann M.A., Distefano P.S., Alnemri E.S., Bertin J.;
 "CARD10 is a novel caspase recruitment domain/membrane-associated
 guanylate kinase family member that interacts with Bcl10 and activates
 NF-kappa B.";
 J. Biol. Chem. 276:21405-21409(2001).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21255663; PubMed=11356195;
 RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschoopp J.;
 "Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10
 phosphorylation and NF-kappaB activation.";
 FEBS Lett. 496:121-127(2001).
 [3]
 RN ERRATUM.
 RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschoopp J.;
 FEBS Lett. 505:198-198(2001).
 [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
 RA Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Garder C., Carter M.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA Mcclay J., McLaren S., McMurray A., Milne S.A., Mortimore B.J.,

RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hua A., Kenton S., Lal H., Leo H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Zhang M., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozerky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA Kim U.J., Shizuya H., Simon M.I., Dumanaki J.P., Peyrard M., Kedra D.,
RA Seroussi E., Frausson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.,
RT "The DNA sequence of human chromosome 22." ;
RL Nature 402:489-495(1999).
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC -!- SUBUNIT: CARD10 and Bcl10 bind to each other by CARD-CARD
CC interaction. They both participate in a complex with MALT1, where
CC MALT1 binds to Bcl10 (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- TISSUE SPECIFICITY: Detected in adult heart, kidney and liver;
CC lower levels in intestine, placenta, muscle and lung. Also found
CC in fetal lung, liver and kidney.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- CAUTION: Supposed to contain a SH3, a PDZ and a guanylate kinase-
CC like domain. But none of these 3 domains are detected by PROSITE,
CC Pfam or SMART.
CC -!- CAUTION: Ref 4 sequence differs from that shown due to various
CC gene identification problems.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY028896; AAK26165.1; -
DR EMBL; AY032928; AAK54454.1; -
DR EMBL; AL049851; CAB63075.1; ALT_SEQ.
DR EMBL; AL049851; CAB63076.1; ALT_SEQ.
DR EMBL; AL022315; CAB42832.1; ALT_SEQ.
DR PROSITE; PS50209; CARD; 1.
KW Coiled coil.
FT DOMAIN 23 115 CARD.
FT DOMAIN 138 456 COILED COIL (POTENTIAL).
FT DOMAIN 567 574 POLY-SER.
FT CONFLICT 289 289 Q -> R (IN REF. 4).
FT CONFLICT 917 917 K -> Q (IN REF. 4; CAB63075).
FT CONFLICT 932 932 R -> L (IN REF. 4).
SQ SEQUENCE 1032 AA; 115946 MW; 8377319AB82A0949 CRC64;
Query Match 22.5%; Score 329.5; DB 1; Length 1032;
Best Local Similarity 31.8%; Pred. No. 7.3e-11;
Matches 100; Conservative 50; Mismatches 115; Indels 49; Gaps 6;
QY 10 LOELNQGKGKGVLLRRCCQLQQLHGLGAEFRAEGLRQQLRQQAQRCQLREDWEAG 62
DB 136 LMTEVRLREARKSQLQRCCQLQARGVLEERAGLEQLRQQAQRCQLREDWEAG 195

QY 63 PHEVLRKDEMLSLSHYSNALQEKELAAASRCRSIQBELYLLKQELQRANMVSSCELELQ 122
DB 196 SLELLRLKDNENYMTARLAQSLSEKNSAVLRSDQLAVDQLKLVSR--LEEBECAL-- 250
QY 123 EQSLRTASDQSGDEENLRKEENE-----KLSR-----LTFSLAE----- 158
DB 251 ---LRRAGPPPPGAAEKEKEKEKEKEDPNVDVLSLRAENQQLTASLREGLQGEASR 307
QY 159 -----KDILEQSLDEARGSRQELVERHISLRRAVAERQRYWEKEBOTLLQF 208
DB 308 PGAGSERILLDIIEHWDREAQDSRQELCQKLKHAQVQELQWAEELRQYLOEMEDRLKH 367
QY 209 QKSKMACOLYREKNAQQAQCEQKRDQAYSARDASQAREISQSLSEKDSLRROFVELT 268
DB 368 RTLQKDCDLYKHRMATVLAQLEETEKEKRDQAIQSRDRIQLQYQSOLIEKQYRKQVRGLE 427
QY 269 DQVCELRTQLRQLQ 282
DB 428 AERDELLTTLTSL 441
RESULT 5
ID CARA_MOUSE STANDARD; PRT; 1021 AA.
AC P58660;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 10 (Bcl10-interacting MAGUK protein
DE 1) (Biml1).
GN CARD10 OR BIMP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=21391892; PubMed=11387339;
RA McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A.,
RA Li Q., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W.,
RA Nunez G.;
RA "Biml1, a MAGUK family member linking protein kinase c activation to
RT Bcl10-mediated NF-kappa B induction.";
RL J. Biol. Chem. 276:30589-30597(2001).
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC -!- SUBUNIT: CARD10 and Bcl10 bind to each other by CARD-CARD
CC interaction. They both participate in a complex with MALT1, where
CC MALT1 binds to Bcl10.
CC -!- TISSUE SPECIFICITY: Highly expressed in kidney, heart followed by
CC brain, lung, liver, skeletal muscle and testis.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- CAUTION: Supposed to contain a SH3, a PDZ and a guanylate kinase-
CC like domain. But none of these 3 domains are detected by PROSITE,
CC Pfam or SMART.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF363456; AAK60136.1; -
DR PROSITE; PS50209; CARD; 1.
KW Coiled coil.
FT DOMAIN 23 115 CARD.
FT DOMAIN 138 450 COILED COIL (POTENTIAL).
FT DOMAIN 558 565 POLY-SER.
SQ SEQUENCE 1021 AA; 114413 MW; 4811A09BD8BF792C CRC64;
Query Match 22.1%; Score 322.5; DB 1; Length 1021;

```
Best Local Similarity 31.1%; Pred. No. 1.7e-10;
Matches 100; Conservative 51; Mismatches 126; Indels 45; Gaps 6;
QY 10 LOEELNQKRGQEVLLRRCCQQLQ-----EHLGLAETRAEGLHQLADHGRMKREVSVA 61
   | : : : : : | | | | | | | | | | | | | | | | | | | | | | | |
Db 136 LMTEVRLREARKSQLHQEQLOAQRALAEERAGL-FQRLREQQAQRCORLREDWEA 194
   | : : : : : | | | | | | | | | | | | | | | | | | | | | | | |
QY 62 HPEVRLRKDEMLSLSHYSNALQKEKLAASRCRSLSQEEYLLKQELQORANMYSCELEL 121
   | : | | | | | : : : : : | | | | | | | | | | | | | | | | |
Db 195 GSLELRLKLDENYMTAMRLAQLSEKNSAVLRSDQLAVDQLKLVSR--LEECCAL-- 250
   | : | | | | | : : : : : | | | | | | | | | | | | | | | | |
QY 122 QQSRLTASDQSGDEE-----LNRLKEEKLRLSFTSLAE----- 158
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 251 ----LRRAGPPPGAEEKEREPDPGADLLSELRAENORLTASLOEQLEOGLQPMSPGAAG 306
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 159 -----KDILEQSLDARGSRQELVERIHSRLRERAAVAERQRYWEEKQETLLQFKSKM 213
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 307 SERILLDILEHDWRQAQDSRQELCKQLHVAQVQELQWAEELDKYLOEMEDLRLKXRTLLK 366
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 214 ACOLYREKVNALQAOVCELOKQERQAYSARDSAQREISOSLVKESLRRQVFEELTDQVCE 273
   | | | : : : | : | : | | | | | | | | | | | | | | | | |
Db 367 DCDLYKXRMATVLAOLEETEKERQAIQSRDRIQLQYSOSLIEKQYRKQVRGLEAERDE 426
   | | | : : : | : | : | | | | | | | | | | | | | | | | |
QY 274 LRTQLRQLQAEPPGVLKQEPART 295
   | | : : | : | | | | | | | | | | | | | | | | | |
Db 427 LTTVTTSLEGTAMLEAQLQRT 448
   | | : : | : | | | | | | | | | | | | | | | | | |

RESULT 6
CAR9_RAT
ID CAR9_RAT STANDARD; PRT; 536 AA.
AC Q9EPY0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 9 (rCARD9).
GN CARD9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=20576268; PubMed=11053425;
RA Bertin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D.,
RA Poyet J.-L., Merriam S., Du M.-Q., Dyer M.J.S., Robison K.E.,
RA DiStefano P.S., Alnemri E.S.;
RT "CARD9 is a novel caspase recruitment domain-containing protein that
RT interacts with Bcl10/CLAP and activates NF-kappa B.";
RL J. Biol. Chem. 275:41082-41086(2000).
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 (By similarity).
CC -!- SUBUNIT: Self-associates. CARD9 and Bcl10 bind to each other by
CC CARD-CARD interaction (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF311288; AAC28791.1; -.
DR InterPro; IPR001315; CARD.
DR PROSITE; PS50209; CARD; 1.
KW Coiled coil.
FT DOMAIN 6 98 CARD.
FT DOMAIN 117 277 COILED COIL (POTENTIAL).
FT DOMAIN 303 420 COILED COIL (POTENTIAL).
SQ SEQUENCE 536 AA; 62631 MW; 6F33089CB7E6BAC9 CRC64;
```

```
Query Match 19.4%; Score 283.5; DB 1; Length 536;
Best Local Similarity 29.4%; Pred. No. 1.1e-08;
Matches 91; Conservative 58; Mismatches 123; Indels 37; Gaps 8;
QY 17 EKGQKEVLLRRCCQQLQEHLL-----GLAETRAEGLHQLADHDS-----RMKREVSVAH 62
   | | : | | : | | : : | | : : : | | | | | | | | | | | |
Db 112 ESGLTQLLMTYMKLQKKVQDILLTALLSSKDDFIKELRVKDSLLRKHOERVQRLKECELS 171
   | : | | | | | : : : : : | | | | | | | | | | | | | | | | |
QY 63 PHEVRLRKDEMLSLSHYSNALQKEKLAASRCRSLSQEEYLLKQELQORANMYSCELELQ 122
   | : | | | | | : : : : : | | | | | | | | | | | | | | | | |
Db 172 SAE LRCKDENYDLAMRLAHLSEKGAALMRNRDLQLEVDQLRHSLSMKAEE--DDCKVERK 229
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 123 EQ-SLRTASDQSGDEELNRLKEENE---KLRSLTFSLAER-----DILEQSLSDE 168
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 230 HTLKL RHAMEQPSQELLWDLQOERDLLQARVQOELEVSVOEGKLRHNSPYIQVLEEDWRQ 289
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 169 ARGSQRELVERIHSRLRERAAVAERQRYWEEKQETLLQFKSKMACQYLRKVNALQAO 228
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 290 ALQEHQEQASTIFSRLKDLRQAEALTRCMEEKEMFELQCLALRKDAKMYKDRIEAILQO 349
   | | | : : : | : | : | | | | | | | | | | | | | | | | |
QY 229 VCELOKQERQOAYSARDSAQREISOSLVKESLRRQVFEELTDVCELRTOL-----RLQQA 283
   | : | | | | | : : : : : | | | | | | | | | | | | | | | | |
Db 350 MEEVSIERDQAMTSREELHAQCAQSFQDKDKLRKQVRELDERADELQQLQFOTESRLAA 409
   | : | | | | | : : : : : | | | | | | | | | | | | | | | | |
QY 284 EPPGVLKQE 292
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 410 E--GRLKQK 416

RESULT 7
CAR9_HUMAN
ID CAR9_HUMAN STANDARD; PRT; 536 AA.
AC Q9H257; Q9H854;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 9 (hCARD9).
GN CARD9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20576268; PubMed=11053425;
RA Bertin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D.,
RA Poyet J.-L., Merriam S., Du M.-Q., Dyer M.J.S., Robison K.E.,
RA DiStefano P.S., Alnemri E.S.;
RT "CARD9 is a novel caspase recruitment domain-containing protein that
RT interacts with Bcl10/CLAP and activates NF-kappa B.";
RL J. Biol. Chem. 275:41082-41086(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Retinoblastoma;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-492 FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Activates NF-kappaB via Bcl10.
CC -!- SUBUNIT: Self-associates. CARD9 and Bcl10 bind to each other by
CC CARD-CARD interaction.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
```

```
CC -!- TISSUE SPECIFICITY: Highly expressed in spleen. Also detected in
CC liver, placenta, lung, peripheral blood leukocytes and in brain.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 360.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF311287; AAC28790.1; -
DR EMBL; AK024001; BAB14766.1; ALT_FRAME.
DR EMBL; BC008877; AAH08877.1; -
DR GenBank; HGNC:16391; CARD9.
DR InterPro; IPR001315; CARD.
DR PROSITE; PS50209; CARD; 1.
KW Coiled coil. 6 98
FT DOMAIN 117 277
FT DOMAIN 332 419
FT DOMAIN 412 492
FT CONFLICT 12 12 N -> S (IN REF. 3).
FT CONFLICT 482 492 LSSGPPPEKER -> PAGLPGIGAVC (IN REF. 3).
FT SEQUENCE 536 AA; 62267 MW; 6EB1835315B83DE5 CRC64;
SQ
Query Match 18.8%; Score 274.5; DB 1; Length 536;
Best Local Similarity 28.4%; Pred. No. 3.2e-08;
Matches 87; Conservative 61; Mismatches 125; Indels 33; Gaps 7;
QY 17 EKGQEVLLRRCOOLEHL-----GLAETRAEGHLQLEADHS-----RMKREVSAAH 62
DB 112 ESGUTQLLTVEYKQKKQYDQLTALLSSKDDFKELRVKDSLLRKHQERVQRKKECEAG 171
QY 63 FHEVRLKDEMLSLSHYNAQELKELAAASCRSLOEELYLLKQELQORANMVSSCELELQ 122
DB 172 SRELKCKEENYDLMLRLAHQSEKGAALMRNRDLQLEIDQLKHSLSMAE--DDCKVERK 229
QY 123 EQ-SLRTPASQESGDEELNRLKEE---NEKLSLTFSLAEK-----DILEQSLDE 168
DB 230 HTLKLHRAHEQPSQELLWELQEKALLQARVOELASVOEGLDRSSPYIOVLEEDWRQ 289
QY 169 ARGSRQELVERIHSRLRERAAERQEQYWEKEQTLLOFQKSKMACOLYREKVNALQ 228
DB 290 ALRDHQEQANTIFSLRKDLRQGEARRLCRCMEKEFELQCLALRKDSKMYKDRIEAILLQ 349
QY 229 VCELQKRDQAYSARDSAQREISQSLVEKDSLRQVFEITDQVCELTQLRLQLOAEPGV 288
DB 350 MEEVAIERDQAIATREELHAHQHARGLEQKDALRKQVRELGEKADELQQLQVFOCEAQ---L 406
QY 289 LKQEAR 294
DB 407 LAVEGR 412
RESULT 8
TRHY_RABIT
ID TRHY_RABIT STANDARD; PRT; 1407 AA.
AC P37709;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trichohyalin.
GN THH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Fietz M.J., Rogers G.E.;
```

```
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -!- SUBUNIT: HOMODIMER (PROBABLE).
CC -!- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FLIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.
CC -!- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -!- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z19092; CAA79519.1; -
DR PIR; S28589; S28589.
DR HSP; P02633; 4ICB.
DR InterPro; IPR001751; CaBP_S100.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 1.
DR Pfam; PF01023; S_100; 1.
DR ProDom; PD003407; CaBP_S100; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS00303; S100_CaBP; 1.
KW Keratinization; Repeat; Calcium-binding.
FT DOMAIN 1 91 S-100 LIKE.
FT CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
FT CA_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
FT SEQUENCE 1407 AA; 183781 MW; AE17D2A159F12B7F CRC64;
SQ
Query Match 13.9%; Score 202.5; DB 1; Length 1407;
Best Local Similarity 27.7%; Pred. No. 0.00057;
Matches 97; Conservative 48; Mismatches 108; Indels 97; Gaps 13;
QY 11 QEELNQEKQKQK---EVLRLRCQQLQEHLGLAETRAEGLHQL-----EADHSRMKRE 58
DB 641 EQELQERERKLREEEQQLLRREEQ-----ELQERERKLREEEQQLLRREEERLRQ 692
QY 59 VSAHFHEVRLKDEMLSLSHYNAQELKELAAASCRSLOEELYLLKQELQORANMVSSCE 118
DB 693 ERA---RKLREEQLL-----RQEEQLRQERERKLREEEQQLLRREEQLRQERDRK 741
QY 119 LELEQSLRTASDQESGDEELNRLKEENKLSLTFSLAEKDLILEQSLDEARGSQE--- 175
DB 742 LREEEQLL-----QESEERLRQEREQQLRRERDRKFRREEQLQEREEERLRQERER 796
QY 176 -----LVERIHSRL-----ERAVAAERQR-----EQWEEK 201
```

Db 797 KLREEEQLLQREERLRRQERKRLREEQQLQREBERLRRQERKRLREEQQLRQE 856

Qy 202 EQLTQFQSKMACOLYREKYNALQAQCEQLQKRD-----QAYSARDSA 246

Db 857 EQELQERARKL-----REEQQLRQEQLRQERDRKRLREEQQLRQEQLRQERDRK 911

Qy 247 QREISQSL--VEKDSLRQVFE--LTDQVCELRQLRQAEPPGVLRQE 292

Db 912 LREEQQLQSEERLRRQERKRLREEQQLRREEQQLRREERARKLREE 961

RESULT 9

CENE_HUMAN STANDARD; PRT; 2663 AA.

AC Q02224;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Centromeric protein E (CENP-E protein).

GN CENPE.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93024922; PubMed=1406971;

RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;

RT "CENP-E is a putative kinetochore motor that accumulates just before

RT mitosis.";

RL Nature 359:536-539(1992).

RN [2]

RP CHARACTERIZATION.

RX MEDLINE=95196755; PubMed=7889940;

RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;

RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed

RT microtubule motor.";

RL EMBO J. 14:918-926(1995).

RN [3]

RP CHARACTERIZATION.

RX MEDLINE=98437347; PubMed=9763420;

RA Chan G.K.T., Schaar B.T., Yen T.J.;

RT "Characterization of the kinetochore binding domain of CENP-E reveals

RT interactions with the kinetochore proteins CENP-F and HUBR1.";

RL J. Cell Biol. 143:49-63(1998).

CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE

CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE

CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT

CC AND/OR SPINDLE ELONGATION.

CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.

CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING

CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS

CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.

CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Z15005; CAA78727.1; -

DR PIR: S28261; S28261.

DR HSSP: P17119; 3RAR.

DR Genew: HGNC:1856; CENPE.

DR MIM: 117143; -

DR InterPro: IPR001752; kinesin_motor.

DR Pfam: PF00225; kinesin.1.

DR PRINTS: PR00380; KINESINHEAVY.

DR SMART: SM00129; KISC.1.

DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.

DR PROSITE: PS0067; KINESIN_MOTOR_DOMAIN2; 1.

KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis.

FT DOMAIN 1 335 KINESIN-MOTOR

FT DOMAIN 336 2471 COILED COIL (POTENTIAL).

FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).

FT NP_BIND 86 93 ATP (BY SIMILARITY).

SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8CB8 CRC64;

Query Match 12.9%; Score 188.5; DB 1; Length 2663;

Best Local Similarity 24.2%; Pred. No. 0.006;

Matches 77; Conservative 50; Mismatches 110; Indels 81; Gaps 12;

Qy 3 LAGATGSLQEEELNKGQKEVLLRRCCQQLQEHGLAETRAEGLHOLEADHSRMKREV--- 59

Db 718 LEGKITDLQKELNKEVEENEALREEVILLSE-----LKSLPSEVERLRKEIQDK 766

Qy 60 SAHFHEVLRKDEMLSLSHYSNALQEKELAAASRCRSQEEYLLYLKQELQRANMYSSCEL 119

Db 767 SEELHIITSEKDLKFSEVH-----KE---SRVQGLLEEIGTKTKDDL----- 805

Qy 120 ELQEOSLTASDOESGDEELNRLK-----EENEKLSLTFSLAEKDLILQES 165

Db 806 -----ATTQSNYKSTDOEFQNFKLHMDFEQYKMWLEENRMNQEIIVLSKE---AQK 856

Qy 166 LDEARGS-----RQELVERIHSRLRERAVAAERQRYWEKEOTLLQFOKSKMACOL 217

Db 857 FDSGLGALKTELSYKTELQEKTRVQERLNEMLQKQL-ENRDSPLQTVREK---TL 912

Qy 218 YREKYNALQAQVCELOKQERDQAYSARDSAQREISOSLVEKDSLRQVFEITDQVCELRTQ 277

Db 913 ITEKLQQTLEEVKTLTQBEKDKLQQLQESLQ-----TERDQLKSDIHDVTVMNIDTQEQ 965

Qy 278 LRQLQAEPPGVLRQKQART 295

Db 966 LRNALES-----LKHQDET 979

RESULT 10

MYHA_RAT STANDARD; PRT; 1976 AA.

AC Q9ULT0;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,

DE type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).

GN MYH10.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Brain;

RX MEDLINE=20483650; PubMed=11027611;

RA Yam J.W.P., Chan K.W., Li N., Hsiao W.L.W.;

RT "Molecular cloning and functional analysis of the promoter region of

RT rat nonmuscle myosin heavy chain-B gene.";

RL Biochem. Biophys. Res. Commun. 276:1203-1209(2000).

CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,

CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND

CC CAPPING.

CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY

CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2

CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).

CC -!- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING

CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,

CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF139055; AAF61445.1; -.
CC HSSP; P10587; 1BR2.
CC InterPro; IPR000048; IQ_region.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC InterPro; IPR003017; Spectrin.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00063; myosin_head; 1.
CC Pfam; PF00612; IQ; 1.
CC Pfam; PF01576; Myosin_tail; 1.
CC Pfam; PF02736; Myosin_N; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC PRODOM; PR000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS50096; IQ; 1.
CC Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Alkylation; Multigene family.
FT DOMAIN 1 785
FT DOMAIN 786 815
FT DOMAIN 845 1976
FT NP_BIND 178 185
FT MOD_RES 701 701 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1976 AA; 228963 MW; E32708BF9BF2B470 CRC64;

Query Match 12.8%; Score 186.5; DB 1; Length 1976;
Best Local Similarity 22.8%; Pred. No. 0.0057;
Matches 79; Conservative 70; Mismatches 115; Indels 83; Gaps 14;

QY 4 AGAIGSLQELN-----QKQGKEVLLRRCCQLQEHLG 36
Db 772 AGVLAHEERDLKTYDIIFFQAVCRGYLARKAKKQQQLSALKVLRNC---AAYLK 828
QY 37 LAF-----TRAGLHQLADHSRMRKREVS AHFHEVLRLKD-----EMLSLSLHYSN 82
Db 829 LRHWQWVRVFTKVKPLQV---TRQEELOAKDELLKVKKEQTKVGELEEMERKHQQ 884
QY 83 ALQEKELASRCRSLQELVLLKQELQRANVYSSCELELQEQSLRTASDOESGDEELNR- 141
Db 885 LLEEKNILAEQLQA-ETELFAEAEAMRARAIAKKQEL---EETLHDLSEKRVGEGERNOI 940
QY 142 LKEENEKRLSLTSLAEKDTLQSLDEARGSRQEL-----VERIHSRLRERAVAAERQR 194
Db 941 LQNEKKKMQ-----AHIQDLEQLDEEGARGKQLQLEKVTAEAKIKKMEEEVLLLEDQN 994
QY 195 EQWKEKEQTLQFQSKMACQLYRKV-----NALQAVQCELO---KERDQAYS 242
Db 995 SKFIKEKK--LMEDRIAECSSQLAEERAKNIAKIRNQEVNIDSELRKKEETROE 1052
QY 243 RDSQAQREISQSLKESDRQVFLTDVCELTQLRQLQAEPPGVL 289
Db 1053 LEXAKRKLKG---ETTDLQDQIAELQAQVDELKVLQTKKEEELQGL 1096

RESULT 11
PLEI_CRIGR
ID PLEI_CRIGR STANDARD; PRT; 4473 AA.
AC Q9J135;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Plectin 1 (PLTN) (PCN) (300-kDa intermediate filament-associated
DE protein) (IFAP300) (Fragment).
GN PLECI.
OS Cricetulus griseus (Chinese hamster).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20334248; PubMed=10873583;
RA Clubb B.H., Chou Y.-H., Hertmann H., Svitekina T.M., Borisy G.G.,
RA Goldman R.D.;
RT "The 300-kDa intermediate filament-associated protein (IFAP300) is a
RT hamster plectin ortholog.";
RL Biochem. Biophys. Res. Commun. 273:183-187(2000).
RN [2]
RP PHOSPHORYLATION.
RX MEDLINE=96215219; PubMed=8626512;
RA Malecz N., Foisner R., Stadler C., Wiche G.;
RT "Identification of plectin as a substrate of p34cdc2 kinase and
RT mapping of a single phosphorylation site.";
RL J. Biol. Chem. 271:8203-8208(1996).
CC -!- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC MICROFILAMENTS. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
CC STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
CC ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC -!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC -!- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
CC -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
CC INTERMEDIATE FILAMENTS DURING MITOSIS.
CC -!- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -!- SIMILARITY: CONTAINS 32 PLECTIN REPEATS.
CC -!- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
CC -!- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF260753; AAF70372.1; -.
CC HSSP; Q01082; 1BKR.
CC InterPro; IPR001589; Actbind_actinin.
CC InterPro; IPR001715; Calponin-like.
CC InterPro; IPR001101; Plectin_repeat.
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF00307; CH; 2.
CC Pfam; PF00681; Plectin; 20.
CC SMART; SM00033; CH; 2.
CC SMART; SM00250; PLEC; 32.
CC SMART; SM00150; SPEC; 4.
CC PROSITE; PS00019; ACTININ_1; PARTIAL.
CC PROSITE; PS00020; ACTININ_2; FALSE_NEG.
CC PROSITE; PS00021; CH; 2.
KW Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
KW Phosphorylation.
FT NON_TER 1 1
FT DOMAIN <1 1259
FT DOMAIN 1260 2544
FT DOMAIN 2545 4473
FT DOMAIN <1 192
FT DOMAIN <1 74
FT DOMAIN 87 189
FT DOMAIN 449 508
FT REPEAT 529 613
FT REPEAT 626 719
FT REPEAT 1104 1204
FT REPEAT 1258 2548
FT REPEAT 2615 2652
```

FT	REPEAT	2653	2690	PLECTIN 2.	
FT	REPEAT	2691	2728	PLECTIN 3.	
FT	REPEAT	2729	2766	PLECTIN 4.	
FT	REPEAT	2770	2804	PLECTIN 5.	
FT	REPEAT	2905	2942	PLECTIN 6.	
FT	REPEAT	2943	2980	PLECTIN 7.	
FT	REPEAT	2981	3018	PLECTIN 8.	
FT	REPEAT	3019	3056	PLECTIN 9.	
FT	REPEAT	3057	3094	PLECTIN 10.	
FT	REPEAT	3274	3311	PLECTIN 11.	
FT	REPEAT	3312	3349	PLECTIN 12.	
FT	REPEAT	3350	3387	PLECTIN 13.	
FT	REPEAT	3388	3425	PLECTIN 14.	
FT	REPEAT	3429	3463	PLECTIN 15.	
FT	REPEAT	3609	3646	PLECTIN 16.	
FT	REPEAT	3647	3684	PLECTIN 17.	
FT	REPEAT	3685	3722	PLECTIN 18.	
FT	REPEAT	3723	3760	PLECTIN 19.	
FT	REPEAT	3760	3797	PLECTIN 20.	
FT	REPEAT	3800	3834	PLECTIN 21.	
FT	REPEAT	3852	3889	PLECTIN 22.	
FT	REPEAT	3890	3927	PLECTIN 23.	
FT	REPEAT	3928	3965	PLECTIN 24.	
FT	REPEAT	3966	4003	PLECTIN 25.	
FT	REPEAT	4007	4041	PLECTIN 26.	
FT	REPEAT	4043	4094	PLECTIN 27.	
FT	REPEAT	4197	4234	PLECTIN 28.	
FT	REPEAT	4235	4272	PLECTIN 29.	
FT	REPEAT	4273	4310	PLECTIN 30.	
FT	REPEAT	4311	4348	PLECTIN 31.	
FT	REPEAT	4349	4386	PLECTIN 32.	
FT	DOMAIN	4039	4089	BINDING TO INTERMEDIATE FILAMENTS (BY SIMILARITY).	
FT	DOMAIN	4414	4429	4 X 4 AA TANDEM REPEATS OF G-S-R-X.	
FT	MOD_RES	4328	4328	PHOSPHORYLATION (BY CDC2).	
SQ	SEQUENCE	4473 AA;	509015 MW;	E144615D361E3484 CRC64;	
Query Match					12.7%; Score 186; DB 1; Length 4473;
Best Local Similarity					27.2%; Pred. No. 0.014;
Matches					87; Conservative 38; Mismatches 137; Indels 58; Gaps 11;
QY	10	LOBELNQKQGVLLRRCQQLQEHGLGTAETRAEGHQLHEADHSRKMREVS AHFEVLR	69		
Db	2192	LEBEAKMQVAEEAARLSVAQAARLRQLABEDLAQQRALAEKMLKEMQAVQEA	2251		
QY	70	KDMLSLSHYSNALQEKELAAASRCSLQELLYLLKQEL--QRANVSVSCELELQEQ---	124		
Db	2252	KAEALLO-----QOKELAQEQARRLOEDKEQMAQQLVEETQGFQRTLEVERQ	2304		
QY	125	-----SLRTAS---DQESGDEELNRLKEE---NEKLSRLTFSLAEKDILEQSLDEAR	170		
Db	2305	SAAERLKLRLMAEMSRQAARAEADAQFRKQAEIGEKLHRTLQATQEKVTLVQTLEIQR	2364		
QY	171	GSQELVERIHSRLPRAVAERQEQYWEKEGTTLLQFKSKMAC-----OLVREKVNALQ	226		
Db	2365	QOQSDHDAER---LREAIAELEREKEL--KQEAQLQLKSEEMQTVOEQIQLQE--TQALQ	2418		
QY	227	AQVCE-----LQKER-----DOAVSARDSAQRETSQSLVEKDSLRROV	264		
Db	2419	KSPFSEKSLQRRERTIEDEKAKLEOLFQDEVAKAQQLEEQQRQQRQMEQEKQLVASM	2478		
QY	265	FELTDQVCELTQLRLOLQAE	284		
Db	2479	EEARRQCEAEAEVRRKQEE	2498		
RESULT 12					
ID	PLE1_HUMAN	STANDARD;	PRT;	4684 AA.	
AC	Q15149; Q15640; Q15148;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				

DE	Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HDI).	
GN	PLEC1.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
OX	[1]	
RP	SEQUENCE FROM N.A. (ISOFORM 1).	
RC	TISSUE=Placenta;	
RX	MEDLINE=96210632; PubMed=8633055;	
RA	Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;	
RT	"Human plectin: organization of the gene, sequence analysis, and	
RT	chromosome localization (8q24).";	
RL	Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).	
RN	[2]	
RP	SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.	
RX	MEDLINE=96312447; PubMed=8698233;	
RA	McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B.,	
RA	Bullrich F., Burgeson R.E., Amano S., Hudson D.L., Owarike K.,	
RA	McGrath J.A., McMillan J.R., Eady R.A.J., Leigh I.M., Christiano A.M.,	
RA	Uitto J.;	
RT	"Loss of plectin causes epidermolysis bullosa with muscular dystrophy:	
RT	cdna cloning and genomic organization.";	
RL	Genes Dev. 10:1724-1735(1996).	
RN	[3]	
RP	VARIANT MD-EBS 1003-GLN--ALA-1005 DEL.	
RX	MEDLINE=97049959; PubMed=8894687;	
RA	Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,	
RA	Hachisuka H., Nishikawa T., McLean W.H.I., Uitto J.;	
RT	"Homozygous deletion mutations in the plectin gene (PLEC1) in patients	
RT	with epidermolysis bullosa simplex associated with late-onset	
RT	muscular dystrophy.";	
RL	Hum. Mol. Genet. 5:1539-1546(1996).	
RN	[4]	
RP	VARIANT MD-EBS LEU-429 INS.	
RX	MEDLINE=21090821; PubMed=11159198;	
RA	Bauer J.W., Rouan F., Kofler B., Rezniczek G.A., Kornacker I.,	
RA	Wische W., Hametner R., Klausegger A., Huber A., Pohla-Gubo G.,	
RA	Wiche G., Uitto J., Hintner H.;	
RT	"A compound heterozygous one amino-acid insertion/nonsense mutation in	
RT	the plectin gene causes epidermolysis bullosa simplex with plectin	
RT	deficiency.";	
RL	Am. J. Pathol. 158:617-625(2001).	
CC	-!- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND	
CC	MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESOMOSOMES OR	
CC	HEMIDESOMOSOMES. COULD ALSO BIND MUSCLE PROTEINS SUCH AS ACTIN TO	
CC	MEMBRANE COMPLEXES IN MUSCLE. MAY BE INVOLVED NOT ONLY IN THE	
CC	FILAMENTS NETWORK, BUT ALSO IN THE REGULATION OF THEIR DYNAMICS.	
CC	-!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.	
CC	-!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE	
CC	PRODUCED BY ALTERNATIVE SPLICING.	
CC	-!- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN	
CC	MUSCLE, HEART, PLACENTA AND SPINAL CORD.	
CC	-!- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH	
CC	VIMENTIN, DESMIN, GAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-	
CC	AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.	
CC	-!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM	
CC	INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).	
CC	-!- DISEASE: DEFECTS IN PLEC1 ARE THE CAUSE OF EPIDERMOLYSIS BULLOSA	
CC	WITH MUSCULAR DYSTROPHY (MD-EBS OR EB-MD), AN AUTOSOMAL RECESSIVE	
CC	DISORDER CHARACTERIZED BY EPIDERMAL BLISTER FORMATION AT THE LEVEL	
CC	OF THE HEMIDESOMOSOME AND ASSOCIATED LATE-ONSET MUSCULAR DYSTROPHY.	
CC	-!- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.	
CC	-!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.	
CC	-!- SIMILARITY: CONTAINS 33 PLECTIN REPEATS.	
CC	-!- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.	
CC	-!- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z54367; CAA91196.1; -
DR EMBL; U53204; AAB05427.1; -
DR EMBL; U63610; AAB05428.1; -
DR EMBL; U63609; AAB05428.1; JOINED.
DR EMBL; X97053; CAA65765.1; -
DR HSSP; Q01082; LBRK.
DR Genew; HGNC:9069; PLEC1.
DR MIM; 601282; -
DR MIM; 226670; -
DR InterPro; IPR001589; Actbind_actinin.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR001101; Plectinin-repeat.
DR InterPro; IPR005326; S10_plectin.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00681; Plectin; 19.
DR Pfam; PF03501; S10_plectin; 1.
DR SMART; SM00033; CH; 2.
DR SMART; SM00250; PLEC; 33.
DR SMART; SM00150; SPEC; 5.
DR PROSITE; PS00019; ACTININ_1; FALSE_NEG.
DR PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR PROSITE; PS50021; CH; 2.
KW Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
KW Phosphorylation; Alternative splicing; Epidermolysis bullosa;
KW Disease mutation.
FT DOMAIN 1 1470 GLOBULAR 1.
FT DOMAIN 1471 2755 CENTRAL FIBROUS ROD DOMAIN.
FT DOMAIN 2756 4684 GLOBULAR 2.
FT DOMAIN 175 400 ACTIN-BINDING.
FT DOMAIN 179 282 CH 1.
FT DOMAIN 295 397 CH 2.
FT REPEAT 645 710 SPECTRIN 1.
FT REPEAT 740 824 SPECTRIN 2.
FT REPEAT 837 930 SPECTRIN 3.
FT REPEAT 1315 1415 SPECTRIN 4.
FT DOMAIN 1469 2756 COILED COIL (POTENTIAL).
FT REPEAT 2826 2863 PLECTIN 1.
FT REPEAT 2864 2901 PLECTIN 2.
FT REPEAT 2902 2939 PLECTIN 3.
FT REPEAT 2940 2977 PLECTIN 4.
FT REPEAT 2981 3015 PLECTIN 5.
FT REPEAT 3116 3153 PLECTIN 6.
FT REPEAT 3154 3191 PLECTIN 7.
FT REPEAT 3192 3229 PLECTIN 8.
FT REPEAT 3230 3267 PLECTIN 9.
FT REPEAT 3268 3305 PLECTIN 10.
FT REPEAT 3306 3343 PLECTIN 11.
FT REPEAT 3485 3522 PLECTIN 12.
FT REPEAT 3523 3560 PLECTIN 13.
FT REPEAT 3561 3598 PLECTIN 14.
FT REPEAT 3599 3636 PLECTIN 15.
FT REPEAT 3640 3674 PLECTIN 16.
FT REPEAT 3820 3857 PLECTIN 17.
FT REPEAT 3858 3895 PLECTIN 18.
FT REPEAT 3896 3933 PLECTIN 19.
FT REPEAT 3934 3971 PLECTIN 20.
FT REPEAT 3975 4008 PLECTIN 21.
FT REPEAT 4063 4100 PLECTIN 22.
FT REPEAT 4101 4138 PLECTIN 23.
FT REPEAT 4139 4176 PLECTIN 24.
FT REPEAT 4177 4214 PLECTIN 25.
FT REPEAT 4218 4252 PLECTIN 26.
FT REPEAT 4265 4305 PLECTIN 27.
FT REPEAT 4319 4356 PLECTIN 28.
FT REPEAT 4408 4445 PLECTIN 29.
FT REPEAT 4446 4483 PLECTIN 30.
FT REPEAT 4484 4521 PLECTIN 31.
FT REPEAT 4522 4559 PLECTIN 32.

FT REPEAT 4560 4597 PLECTIN 33.
FT DOMAIN 4250 4300 BINDING TO INTERMEDIATE FILAMENTS (BY
FT SIMILARITY).
FT 4 X 4 AA TANDEM REPEATS OF G-S-R-X.
FT MOD_RES 4625 4640 PHOSPHORYLATION (BY CDC2) (BY
FT SIMILARITY).
FT VARSPLIC 1 174 MVAGMLMPDRDLRAIYEVLFREGVMVAKDRRPSLHPHP
FT GVTNLQVMRAMASLRARGLVRETFAMCHFWYLTNEGIAHL
FT ROYLHLPEIIVASLQVRVRPVAMVMPARRTPHVQAVQGPL
FT GSPKRGPLPTEQRLYRRKELEVSPETVPVPATTORTLA
FT RGPPEPAPAT -> MSGEDAERVAVSDVNSGSGSPGCD
FT TLPWNLGKTORSRRSGGAGNSGVLDPAERAVIRIA (IN
FT ISOFORM 2 AND ISOFORM 3).
FT MISSING (IN ISOFORM 3).
FT L -> LL (IN MD-EBS).
FT /FTid=VAR_011336.
FT MISSING (IN MD-EBS).
FT /FTid=VAR_011337.
FT S -> F (IN REF. 2).
FT N -> D (IN REF. 2).
FT H -> N (IN REF. 2).
FT I -> V (IN REF. 2).
FT Q -> R (IN REF. 2).
FT N -> Y (IN REF. 2).
FT V -> A (IN REF. 2).
FT V -> L (IN REF. 2).
FT L -> V (IN REF. 2).
FT V -> L (IN REF. 2).
FT I -> M (IN REF. 2).
FT A -> T (IN REF. 2).
FT WLC -> RLR (IN REF. 2).
FT Q -> E (IN REF. 2).
FT L -> A (IN REF. 2).
FT K -> R (IN REF. 2).
FT K -> N (IN REF. 2).
FT S -> R (IN REF. 2).
FT R -> Q (IN REF. 2).
FT S -> A (IN REF. 2).
FT K -> E (IN REF. 2).
FT E -> A (IN REF. 2).
FT F -> L (IN REF. 2).
FT F -> L (IN REF. 2).
FT S -> A (IN REF. 2).
FT G -> A (IN REF. 2).
FT R -> S (IN REF. 2).
FT K -> Q (IN REF. 2).
FT E -> Q (IN REF. 2).
FT N -> H (IN REF. 2).
FT V -> A (IN REF. 2).

Query Match 12.7%; Score 185.5; DB 1; Length 4684;

Best Local Similarity 27.0%; Pred. NO. 0.015;

Matches 88; Conservative 41; Mismatches 134; Indels 63; Gaps 11;

QY 10 LOEELNQEKGQEVLLRRCQQLQOEHLGLAETRAEGLHLEADHSRMKREVSAAHFVLRRL 69
||||| : | | | | | : | | | | | : | | | | | : | | | | |

Db 2403 LOEEAEKMKQVAEEAARLSVAQAEEAARLRLQAEEDLAQOALAEKMLKEKMQAVQEAETRL 2462
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 70 KDEMLSLSHYSNALQEKELAAASRCRSLOEELYLLKQEL--QRANWVSSCELELEQ--- 124
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 2463 KAAEALLQ-----QQELAQEQARRLQEDKEQMAQQLAAETQGFORTLEAERQRLQM 2515
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 125 -----SLRTAS---DQESGDEELNRUKEE-----NEKLRSLTFFSLAEKDILEQSLDEAR 170
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 2516 SAAEAERLKLVAEMSAQAARAEEDAQRFRKQAEIGEKLHRTLATQEKVTLVQTLEIQR 2575
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 171 GSRQELVERIHSRLERAVAAERQRYWEEKETLLQFOKSKMAC-----OLYREKVNALQ 226
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 2576 QOSDHDAAER---LREAIAELEREKEKL--QQEAKLLQLKSEEMQTVQOQLLQETQALQ 2630
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 227 AOYCE-----LQKER-----DOAYSARDSAQREISQSILVEKDSL----- 260
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 2631 SFLSEKDSLLQREREFTEQEKAKLEQLFQDEAVAKAQQLREEQQRQOQOQMEQRQRLVASME 2690
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Query Match	12.7%;	Score 185;	DB 1;	Length 1597;
Best Local Similarity	26.5%;	Pred. No. 0.0055;		

Db	1039	LEERLKEETROE	LEKARKKIDG	ETDLDQOIAELQ	OIDELKIQVAKKEE	LOQALAR	1098
			: :: :	:: :	:		
Qy	222	-----VNALQAOYCE	LKQDRAOYSAQREIS	SOSLVERKDSLR	ROVFEL	267	
Db	1099	GDDETHKNNAL	KVVRELQAOIAELQ	EDFESEKASRNKAEQ	KRDLSELEALTE	LEDT	1158
Qy	268	TDQVC	---ELRTOLRO	LAEP	PGVLKQ	EART	295
Db	1159	LDTTAAQOEL	RTKREQV	AEALKEA	LEETKS	1189	

Search completed: January 22, 2003, 08:53:32
Job time : 19.4987 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:36 ; Search time 45.9946 Seconds
(without alignments)
1321.544 Million cell updates/sec

Title: US-09-767-215-2_COPY_126_420
Perfect score: 1462
Sequence: 1 ELAGAIGSLQELNQEKGQ.....TQLRLQAEPGGVLKQEPART 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	DB ID	Description
1	396.5	27.1	1171	4	Q8TES3	Q8tes3 homo sapien
2	198.5	13.6	691	11	Q8R443	Q8r443 rattus norv
3	196	13.4	612	4	Q96JW2	Q96jw2 homo sapien
4	195.5	13.4	1313	4	Q75033	Q75033 homo sapien
5	194	13.3	2101	4	Q14981	Q14981 homo sapien
6	194	13.3	2115	4	Q14980	Q14980 homo sapien
7	193.5	13.2	691	11	Q9D935	Q9d935 mus musculu
8	193.5	13.2	2238	11	Q70365	Q70365 mus musculu
9	192	13.1	1456	5	Q9V587	Q9v587 drosophila
10	192	13.1	2168	11	Q61043	Q61043 mus musculu
11	191.5	13.1	1690	5	Q9VJ85	Q9vj85 drosophila
12	191.5	13.1	1690	5	Q44929	Q44929 drosophila
13	191	13.1	585	4	Q9UQ07	Q9uq07 homo sapien
14	191	13.1	731	4	Q95287	Q95287 homo sapien
15	191	13.1	731	4	Q8TBA6	Q8tba6 homo sapien
16	189.5	13.0	985	5	Q9U0S5	Q9u0s5 mytilus gal

17	189.5	13.0	1705	5	Q9U0S7	Q9u0s7 mytilus gal
18	189	12.9	729	11	Q9QYE6	Q9qye6 mus musculu
19	188	12.9	729	11	Q88317	Q88317 mus musculu
20	188	12.9	1133	5	Q21022	Q21022 caenorhabdi
21	187.5	12.8	1256	5	Q9XVH4	Q9xvh4 caenorhabdi
22	187	12.8	876	17	Q8TXI4	Q8txi4 methanopyru
23	186.5	12.8	1335	13	Q9PW73	Q9pw73 xenopus lae
24	186.5	12.8	2442	4	O14812	O14812 homo sapien
25	186	12.7	2442	4	O60588	O60588 homo sapien
26	186	12.7	2442	4	O9H450	O9h450 homo sapien
27	185	12.7	645	4	Q9P122	Q9p122 homo sapien
28	185	12.7	670	4	Q96CN5	Q96cn5 homo sapien
29	185	12.7	690	4	Q96J03	Q96ju3 homo sapien
30	185	12.7	691	4	Q9H090	Q9h090 homo sapien
31	185	12.7	1641	11	O88528	O88528 mus musculu
32	185	12.7	2035	11	Q9JMH9	Q9jmh9 mus musculu
33	185	12.7	2055	11	O88938	O88938 mus musculu
34	185	12.7	3187	11	O63714	O63714 rattus norv
35	184.5	12.6	1999	11	O63731	O63731 rattus norv
36	184	12.6	746	5	Q25561	Q25561 naegleria f
37	183	12.5	1618	11	Q9QX19	Q9qx19 rattus norv
38	182.5	12.5	574	11	O8R1D2	O8r1d2 mus musculu
39	182.5	12.5	924	5	O15738	O15738 dictyosteli
40	182.5	12.5	1286	5	Q21025	Q21025 caenorhabdi
41	182	12.4	2354	5	Q9NKT9	Q9nkt9 leishmania
42	181	12.4	1370	13	O73732	O73732 xenopus lae
43	181	12.4	2611	11	Q91ZU8	Q91zu8 mus musculu
44	181	12.4	2954	13	O42263	O42263 xenopus lae
45	180.5	12.3	1927	5	Q25142	Q25142 halocynthia

ALIGNMENTS

RESULT 1

Q8TES3 ID Q8TES3 PRELIMINARY; PRT; 1171 AA.
AC Q8TES3;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE FLJ00120 protein (fragment).
GN FLJ00120.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA Ohara O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human spleen."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074049; BAB84875.1; -.
FT NON_TER 1
SQ SEQUENCE 1171 AA; 134966 MW; FA567ABBC8A703FF CRC64;

Query Match	27.1%	Score	396.5;	DB	4;	Length	1171;
Best Local Similarity	34.0%;	Pred. No.	3.9e-14;				
Matches	102;	Conservative	66;	Mismatches	109;	Indels	23;
Gaps	6;						
QY	3	L	A	G	A	G	T
DB	148	L	M	N	E	V	I
QY	62	H	F	E	V	E	L
DB	207	Y	N	D	E	L	V
QY	122	Q	E	O	S	R	T
DB	264	R	N	O	S	L	K

```

QY 164 QSLDEARGSRQELVERIHSLRERAVAAERQEQYWEKEQTLLQFQSKMACOLYREKVN 223
DB 324 HDRKEALDRQELVRIYINRYIQAQAEELKQELKEEDLELCKSTLGKDCMKYKRMN 383
QY 224 ALQAOVCLEQERQOAYGARDSAQRSEISQSLVEKDSLRQVFEITDQVCELRTQLRQLOA 283
DB 384 TVMLQLEEVERDOAFHSRDEAQTQYSCQLIENDKYRKQIRELEENKNDRIEMVRREA 443

RESULT 2
Q8R443 PRELIMINARY; PRT; 691 AA.
AC Q8R443;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE KIAA1536 protein.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OFA-SD; TISSUE=CEREBELLUM;
RA Vie-Luton M.-P., Francon J.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY078385; AAL05572.1; -.
SQ SEQUENCE 691 AA; 77276 MW; E4FF5AA1784377FE CRC64;

Query Match 13.6%; Score 198.5; DB 11; Length 691;
Best Local Similarity 23.7%; Pred. No. 0.0015;
Matches 88; Conservative 64; Mismatches 119; Indels 101; Gaps 15;

QY 10 LOEELNQEKQKEVLLR-----RCQQL-----OEHLGLAE-----TRA 42
DB 149 LQNLQDSQQRNDLQKLQLEDQVTELSRVQLEAALATARGESLSEQYKGLSRS 208
QY 43 EGLHOLEADHSRMRKRVSAHFHEVLRLKDEMLSLSHYSNALQEKELAAASRCRSL----- 97
DB 209 HG--ELSEERDLSQQGGHVARIIELEDDIQTM-----SDKVLMEVELDRVDMVKALT 262
QY 98 --OEELYLKQLQRANMYSSEL-----ELQE-----QSLRTASQESGDE-- 137
DB 263 REQELKLGQKFEQADKEQSEAELOVTVREENCHLNTLQEAQGRQEEQAGVQVRLKDKVA 322
QY 138 -----ELNRLKEENEKRLSLTFSLAEK-DILEQSLDEARGSRQELVERIH 182
DB 323 HMKDTLGTQYQKVAEPLKEQLRGVQELAAASSQKRAALLGBELASAGARDRTIAELH- 381
QY 183 LRERAVAAE-----RQEQYWEKEQTLLQFQSKMACOLYREKVNLAQAQVC 230
DB 382 -RSRLEVAEVRGLAELSLHMKKEKCQWSKERTGLLOSMEAE-----KDKILKLSAETL 434
QY 231 ELQK-----ERQAYARDSAQRISQSLVEKDSLRQVFEITDQVCELRTQLRQLOA 286
DB 435 RLEKTVQEQERSQSHMFKELTAREKDSLSVLQSEKRELTLSALRVLQKEQELQTERQ 494
QY 287 GVL-----KQAP 294
DB 495 ELLEYMRKLEAR 506

RESULT 3
Q96JV2 PRELIMINARY; PRT; 612 AA.
AC Q96JV2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CDNA FLJ14957 fis. clone PLACE4000009, weakly similar to myosin heavy
DE chain, nonmuscle type B.
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027863; BAB55415.1; -.
DR InterPro; IPR002928; Myosin tail.
DR Pfam; PF01576; Myosin tail; 1.
SQ SEQUENCE 612 AA; 72111 MW; EE0EACB9710B594E CRC64;

Query Match 13.4%; Score 196; DB 4; Length 612;
Best Local Similarity 25.8%; Pred. No. 0.0018;
Matches 74; Conservative 62; Mismatches 101; Indels 50; Gaps 12;

QY 10 LOEELNQEKQKEVLLRRCQQLQEHGLAETRAEGLHOLEADHSRMRKRVSAHFHEVLRL 69
DB 137 LOEENKLGQRSEELRRVAQLRQ-----IEDLKGDEAKAKETLKKYGEIRQL 186
QY 70 KDEMLSLSHYSNALQEKELAAASRCRSLQEEYLLKQLQRANMVSCELELQ-EQSLRT 128
DB 187 EEALVHV-----RKEEKAVSAR-RALENEL-----EAAQGNLSQITQEQKQLSEK 234
QY 129 ASDQESGDEELNRLKEENEKRLSLTFSLAEK-----DILEQSLDEARGSRQELVERIH 182
DB 235 ESEQK---EQLRRLKNEMENRWHLGTIEKLOKEMADIVEAS---RTSTLELQNOLDE 287
QY 183 LRE--RAVAERQEQYWEKEQTLLQFQSKMACOLYREKVNLAQAQVCLEQERQOAY 240
DB 288 YKENRRELAEQMORL-----KEKT-LEAEKSLRTAMKQDGMRLMEELRYQRAQDEAL 342
QY 241 SARSAQREISQSLVE-----KDSLRQVFEITDQVCELRTQLRQ 280
DB 343 TKRQLLEQLTLDLEYLEAKSHLKDSDRSRLVKQMEDKVSQLEMELEE 389

RESULT 4
Q75033 PRELIMINARY; PRT; 1313 AA.
AC Q75033;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE KIAA0445 protein (DJ37C10.5) (KIAA0445).
GN KIAA0445 OR DJ37C10.5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98116662; PubMed=9455484;
RA Seki N., Ohira M., Nagase T., Ishikawa K., Miyajima N., Nakaajima D.,
RA Nomura N., Ohara O.;
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain.";
RL DNA Res. 4:345-349(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Heath P.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007914; BAA32290.1; -.
DR EMBL; AL049569; CAB96825.1; -.
SQ SEQUENCE 1313 AA; 148254 MW; 00DC42FE419EED1 CRC64;

```

```
Query Match      13.4%; Score 195.5; DB 4; Length 1313;
Best Local Similarity 25.9%; Pred. No. 0.0042;
Matches 83; Conservative 56; Mismatches 102; Indels 79; Gaps 12;

Qy 1 ECLAGTSGTQELNQEGKQEVLLRRCQQL-----QEHGLG-A 38
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 210 EALGSLFEVQRLAQLEARREQLEAGQALLAKETLTGELAGLRQOIATQEKASLDK 269
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 39 ETRAEGHLQLEAD--HSRMKREVSAAHFHVLRLKDKMLSLSLHYSNALQEKELAAASRCRSL 97
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 270 ELMAQKLVAEREQAASLRQEAHAEDQLRLQ-----REKEA---WREL 312
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 98 QEELYLKQELQELOR--ANNVSCSELELEQSLRTASDQESGDEELNRLKEENKRLSRTFS 155
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 313 EAERAQSLQSLQEQEELRLARLEAEKEPSELSEIALAQOERDEGL--LIAESEKQALSLK 370
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 156 LAERDILEOSDEARGSRQELVERIHSLRERAAVAAERQEQYWEKEQTLLOFQKSKMAC 215
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 371 ESEKTAISEKL--MGTR-----HSLATISLEMEKQKRDQAQRQED----- 409
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 216 QLYREKVNALQAELOKQERDQAYSARDSAQREISQSL-----VEKDSLRRQVFEITDQV 271
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 410 ---RSTVNALTSFLDLRAQREEAHAHAQEVRLQEQARDLGKQDSCSLREAE----- 460
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 272 CELRTQLRLQAEPPGVLUK 291
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 461 -ELRTQLRLLEDARDGLRRE 479
| | | | | : | | | : | | | | | | | | | | | | | | | |

RESULT 5
ID Q14981 PRELIMINARY; PRT; 2101 AA.
AC Q14981;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE NuMA protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92176238; PubMed=1541636;
RA Compton D.A., Szilak I., Cleveland D.W.;
RT "Primary Structure of NuMA, an Intracellular protein that defines a
RT Novel Pathway for Segregation of Proteins at Mitosis.";
RL J. Cell Biol. 116:1395-1408(1992).
DR EMBL; Z11584; CAA77670.1; -.
SQ SEQUENCE 2101 AA; 236297 MW; 2097F679D4E38E69 CRC64;

Query Match      13.3%; Score 194; DB 4; Length 2101;
Best Local Similarity 26.6%; Pred. No. 0.008;
Matches 93; Conservative 49; Mismatches 104; Indels 104; Gaps 15;

Qy 10 LOEELN---QEK-----GQEVLLRRCQQLQEHGLAETRAEGLHQLAEADHSRMKREVSAAH 62
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 362 LEKLSAALQDKKCLEEKNEILQGLKSLSEHLS-----QLQDNPPQKEGEVLGD 411
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 63 FHEVLRKDEMLSLS-----LHYSNALQEKELAAASRCRSLOEELYL----- 103
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 412 VLQLETLKQEAATLAANNLTQARVEMLETERGQGEAKLLAERGHFEFEKQOQLSLITDL 471
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 104 -----LQELQRANWVSSCELELEQESLRT-----ASDQESGDEELNRLKEE--- 145
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 472 QSSISNLQAEKEELQASQAHCARLTAQVSLTSELTTLNATIQQQ--DOELAGLQQAQKE 530
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 146 -----NEKRLSLTSLAEKDILOESLDEARGSRQELVERIHSLRER 186
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 531 KQAOQAQTLQOQEQASQGLRHQVEQLSSSLKQK---EQLKEV-AEKQEAATQDHA-QQL 585
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 187 AVAARQEQYWEKEQTLLOFQKSKMACQLYREKVNALQAEVCELQERDQAYSARDSA 246
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 586 ATAAE-BREASLRERDAALKOLE-----ALEKEKAAKLEI-----LQOOLQVANEARDSA 634
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 247 QREISQSLVEKDSLRRQVFEI-----TDQVCELRTQLRLQ 282
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 635 QTSVTAQAREKAELSRKVEELQACVETARQEQHEAQVAQVAELEQLRSEQ 684
| | | | | : | | | : | | | | | | | | | | | | | | | |

RESULT 7
ID Q9D935 PRELIMINARY; PRT; 691 AA.
AC Q9D935;
```

```
Db 586 ATAAE-BREASLRERDAALKOLE-----ALEKEKAAKLEI-----LQOOLQVANEARDSA 634
Qy 247 QREISQSLVEKDSLRRQVFEI-----TDQVCELRTQLRLQ 282
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 635 QTSVTAQAREKAELSRKVEELQACVETARQEQHEAQVAQVAELEQLRSEQ 684
| | | | | : | | | : | | | | | | | | | | | | | | | |

RESULT 6
ID Q14980 PRELIMINARY; PRT; 2115 AA.
AC Q14980;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE NuMA protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92176231; PubMed=1541630;
RA Yang C.H., Lambie E.J., Snyder M.;
RT "NuMA: an unusually long coiled-coil related protein in the mammalian
RT nucleus.";
RL J. Cell Biol. 116:1303-1317(1992).
DR EMBL; Z11583; CAA77669.1; -.
SQ SEQUENCE 2115 AA; 238274 MW; DA5686215054DBEE CRC64;

Query Match      13.3%; Score 194; DB 4; Length 2115;
Best Local Similarity 26.6%; Pred. No. 0.0081;
Matches 93; Conservative 49; Mismatches 104; Indels 104; Gaps 15;

Qy 10 LOEELN---QEK-----GQEVLLRRCQQLQEHGLAETRAEGLHQLAEADHSRMKREVSAAH 62
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 362 LEKLSAALQDKKCLEEKNEILQGLKSLSEHLS-----QLQDNPPQKEGEVLGD 411
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 63 FHEVLRKDEMLSLS-----LHYSNALQEKELAAASRCRSLOEELYL----- 103
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 412 VLQLETLKQEAATLAANNLTQARVEMLETERGQGEAKLLAERGHFEFEKQOQLSLITDL 471
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 104 -----LQELQRANWVSSCELELEQESLRT-----ASDQESGDEELNRLKEE--- 145
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 472 QSSISNLQAEKEELQASQAHCARLTAQVSLTSELTTLNATIQQQ--DOELAGLQQAQKE 530
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 146 -----NEKRLSLTSLAEKDILOESLDEARGSRQELVERIHSLRER 186
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 531 KQAOQAQTLQOQEQASQGLRHQVEQLSSSLKQK---EQLKEV-AEKQEAATQDHA-QQL 585
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 187 AVAARQEQYWEKEQTLLOFQKSKMACQLYREKVNALQAEVCELQERDQAYSARDSA 246
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 586 ATAAE-BREASLRERDAALKOLE-----ALEKEKAAKLEI-----LQOOLQVANEARDSA 634
| | | | | : | | | : | | | | | | | | | | | | | | | |
Qy 247 QREISQSLVEKDSLRRQVFEI-----TDQVCELRTQLRLQ 282
| | | | | : | | | : | | | | | | | | | | | | | | | |
Db 635 QTSVTAQAREKAELSRKVEELQACVETARQEQHEAQVAQVAELEQLRSEQ 684
| | | | | : | | | : | | | | | | | | | | | | | | | |

RESULT 7
ID Q9D935 PRELIMINARY; PRT; 691 AA.
AC Q9D935;
```

```
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE 1810009B06R1K protein.
GN 1810009B06R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217831;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Anon H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Sasaki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
DR Nature 409:685-690(2001).
DR EMBL; AK007393; BAB25009.1; -.
DR MGD; MGI:1914738; 1810009B06R1K.
SQ SEQUENCE 691 AA; 77280 MW; 4759E5478839D9B7 CRC64;

Query Match 13.2%; Score 193.5; DB 11; Length 691;
Best Local Similarity 22.9%; Pred. No. 0.0028;
Matches 86; Conservative 66; Mismatches 116; Indels 107; Gaps 16;

QY 10 LQELNQEKGQKEVLLR-----RCQOL-----QEHGLGAE-----TRA 42
Db 149 LQNLDESQQRNLDMLQLQLEDOVTELSRVQLEAALATARQEHSELTEQYKGLSRS 208
QY 43 EQLHQLADHDSRMKREVSAPHFVRLRDEMLSLSHYSNALQEKELAAASCRS-----L 97
Db 209 HG--ELSEERDILSQQGDHVARILELEDDIQTM-----SDKVLKREVELDRVDTVKALT 262
QY 98 QEELYLLKQ-----ELQRANVSSC-ELELQEQSLRTASDQESGDE----- 137
Db 263 REQEKLLRLQKEFQADKEQSEAELOTVRENCNLTNLEEAASR---QEEQAGQVORLKD 319
QY 138 -----ELNRLKENEKRLSTFTSLAEK-DILEQSLDEARSGROELVER 179
Db 320 KLAHMKDITLGAQKQKVAEPLKQLRQVQLEAASSQQAALLGELASAGARDRTTAE 379
QY 180 IHSRLRRAVAE-----BQREYWEKEQTLLOFOKSKMACOLYREKVNALQA 227
Db 380 LH--RSRLEVAEYNGRLAELSLSHMKEKQCSKRTGLQSQWEAE-----KDKILKLSA 431
QY 228 QVCLELQK-----ERDAQSARDSAQREISQSLRQVLFELTDQVCELTRLQRLQA 283
Db 432 EILRLKTEQVEERTQSHVFKTELAREKDSLSVLQSEKRELTELRSALRVLOKEKEFLQT 491
QY 284 EPPGVVL-----KQEAR 294
Db 492 EKQELLEYMRKLEAR 506

RESULT 8
O70365
ID O70365 PRELIMINARY; PRT; 2238 AA.
```

```
AC O70365;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Golgi autoantigen golgin subtype a4.
GN GOLGA4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=BRIN;
RA Cowan D.A., Gay D., Beiler B.M., Zhao H., Yoshino A., Davis J.G.,
RA Tomayko M.M., Murali R., Greene M.I., Marks M.S.;
RT "Characterization of mouse tgolgin-1 (golgin-245/ trans golgi
RT p230/256kd golgin) and its upregulation during oligodendrocyte
RT development.";
RL DNA Cell Biol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=BRIN;
RA Gay D., Beiler B.M., Zhao H., Bhandoola A., Tomayko M.M., Murali R.,
RA Marks M.S., Greene M.I.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=BRIN;
RA Cowan D.A., Gay D., Beiler B.M., Zhao H., Yoshino A., Davis J.G.,
RA Tomayko M.M., Murali R., Greene M.I., Marks M.S.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF051357; AAC05573.2; -.
SQ SEQUENCE 2238 AA; 257562 MW; 494EA2C111F0165B CRC64;

Query Match 13.2%; Score 193.5; DB 11; Length 2238;
Best Local Similarity 24.9%; Pred. No. 0.0091;
Matches 100; Conservative 59; Mismatches 121; Indels 121; Gaps 20;

QY 9 SLOBELNQERKQ-----KEVLLRRQ-----QLQEHGLGAEATRAEGLHQ----- 47
Db 373 NLIQLEQDQKGMVITETKRMQLETLELKEDEIAQLRSHIRKQMTTQGEELREQEKESRAA 432
QY 48 -----LEADHSRMKREVSAPHF-----HEVLRLLKDEMLSL- 77
Db 433 FEELEKALSTAQKTEDAQRMMKMEMDEQMKAVERASEERLRLOHLELSRVQEAASWAKK 492
QY 78 -----LHYSN-ALQEKELA---ASRCRSLOBEL-----YL-LKQELQRA 111
Db 493 NSEQVAAALQKLHAEELASKEQELSRRLREARELOQMRLALEKSRSEYVLLKLTQKEQQ 552
QY 112 NMVSSCELELQEQSLRTASD---QESGDE-----EL-----NRLKEENKRLSLTF 154
Db 553 ESLALELELQKKAILETSENKLOELQGEAEAYRTRILETSLKLSQESKTQSEHLAV 612
QY 155 SL-AEKD-----ILQSLDEARSGROE---LVERIHSL-BERAVAAERQREYWE 200
Db 613 HLEAEKHNKHNKELTALAEQHRTEVEGLOQSDLSLWTRQLQSLSQHQHAAVEELREKTKQE 672
QY 201 KEQTLLOFOKSKMACOLY-----REKVNALQAVCEQLQERDQAYSARDSAQREISQSL 254
Db 673 KD-ALLKEKESLFQAHIDNMNEKTLEKDKQMELESVSSSEALRARDQLABELSVLRL 731
QY 255 VEKSLRQVLFELTDQVCELTRLQRLQAEPGPVLLKQEAR 295
Db 732 GDADKM-KQALE-----AELEEQRRHHQREVGSISEQQEELT 766

RESULT 9
Q9V587
ID Q9V587 PRELIMINARY; PRT; 1456 AA.
AC Q9V587;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
```

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CG1931 protein (Fragment).
GN CG1931.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Adayani A., An H.-J., Andrews-Frankoch C., Miklos G.L.G.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Baldwin D.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Beasley E.M.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy L., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AF003833; AAF58930.1; -.
DR FlyBase; FBgn0033420; CG1931.
DR InterPro; IPR000533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYSIN.
FT NON_TER 1456 1456
SQ SEQUENCE 1456 AA; 164239 MW; 49E04C6716443ECA CRC64;

Query Match 13.1%; Score 192; DB 5; Length 1456;
Best Local Similarity 25.1%; Pred. No. 0.0072;
Matches 87; Conservative 48; Mismatches 116; Indels 96; Gaps 11;

QY 5 GATGSLQEEELNOEKGQEVLLRRCCQQLQEHGLAETRAEGLHQLEADHS 64
DB 170 GALTSLLEEAI---GDKE---KQMAQLRQDRRAE-----HEKQERDLHEREVADYKI 216
QY 65 EVLRKDEMLSLSHYSNALQEKELAAASRCRSLOEELYLLKELQANMVSSCE-----118
DB 217 KRAAESEVEKQLQTRLEAVTERLEIKLEASQSELGSKAELEKA-----TCMGSSA 272
QY 119 -----LELQFQSLR-----127
DB 273 DWESTKQRTARLENERLKHDLERSQTTGRTTMTTSQELDRAQERADKASAEALRTQA 332

QY 128 ----TASDQESGDEELNRLKEENKELRSUTFTSLAEKIDLEQSLDARGSRQEL-----V 177
DB 333 ELRVTQSDAERAREAAALQEKLSQGEVYRLKAK--LENAQGEQESLRQLEKAQSGV 390
QY 178 ERTHSLREAVARAERQEQWYEKEOTLLQFQKSKMACQLYREKYNALQAQVCELOKERD 237
DB 391 SRHADRDRAFS---EVEIKEMEERTQTGLKS---QLQHEK---LQNSLDKAQNEVD 440
QY 238 QAYSARDSAQREISQSLVEKSLRQVVELTDQVCELRQLRQLAQ 284
DB 441 HLQDKLDKACTENRRVLVEKELTYDYNLQSLDRLKALQAARMQKE 487

RESULT 10
Q61043 PRELIMINARY; PRT; 2168 AA.
AC Q61043;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Ninein.
GN NIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Bouckson-Castang V., Moudjou M., Ferguson D.J.P., Mucklow M.,
RA Belkaid Y., Milon G., Crocker P.R.;
RT "Molecular characterisation of ninein, a new coiled-coil protein of
RT the centrosome."
RL J. Cell Sci. 108:0-0(1996).
DR EMBL; U40342; AAR83234.1; -.
DR MGD; MGI:105108; Nin.
DR InterPro; IPR002048; EF-hand.
SQ SEQUENCE 2168 AA; 249168 MW; FDB8EC0F240E58E7 CRC64;

Query Match 13.1%; Score 192; DB 11; Length 2168;
Best Local Similarity 24.4%; Pred. No. 0.011;
Matches 84; Conservative 68; Mismatches 112; Indels 80; Gaps 16;

QY 6 AIGSLQ-EELNOEKGQEVLLRRCCQQLQEHGLAETRAEGL-----HQLEADHS 53
DB 1561 SISNLKLEELN---GQEEELWQKIEIEQEKASIQTWKELKKQVSDLKTKNQLDSENI 1617
QY 54 RMKREYSAHFHVLRUKDEMLSLSHYSNAL-QEKELAAASRCRSLOEELYLLKQEL---- 108
DB 1618 ELSQKNQN-----KEELKTLNQLAEMLCQREEPGACTSEKWEQNASLKEELDHYK 1670
QY 109 -QRANVSSCELEQEQSLRT-ASDOES--GDEELNRLKE-----E 145
DB 1671 VQTSTLVSSLEAELSRIKLQTHVMEGELLKDELRLKHCRLDPLSLDQKMSVLSY 1730
QY 146 NEKLSRTSLAEKIDLEQSLDARGSRQELVERIHSLSRERAVAAERQEQWYEKEQTL 205
DB 1731 NEKL-----LKEEVL---BELKSCADKLAE--SSLLEHRTATMKQETANEQSESL 1779
QY 206 -LQFQKSKMACQLYREKYNALQAQVCELOKERDQAYSARDSAQREIS-----QSLVEKD 258
DB 1780 KSQALVSAQKVDLEDVLQNVNLQMAEIESDLQVTRQKEAVQKQVMSLHROLQNAIDKD 1839
QY 259 SURROQVELTDQVCELRQLRQ-----LQAEPPGVKQEAR 294
DB 1840 ----WVSETAPHLGURGOORRLSWDKLDHLMNEEPOLLCQESK 1879

RESULT 11
Q9VJES PRELIMINARY; PRT; 1690 AA.
ID Q9VJES
AC Q9VJES;
DT 01-MAY-2000 (Tremblrel. 13, Created)

DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)	109	----	ORANVSSCELEL----	QEQSLRTASQESGDEELNRLKE--	ENEKL-RSLTFSL--	156
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)	664	EGTEKSTLLEKTEKELVQSKEQAATLN	DKNQLEKEISDLKQLAEQKLVREMTNAIN	723		
DE	CLIP-190 protein.						
GN	CLIP-190 OR CG5020.						
OS	Drosophila melanogaster (Fruit fly).	157	---	AEKDILEQSLD-----	EARGSRQ--	ELVERIHSLR----	184
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;						
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;						
OC	Ephydroidea; Drosophilidae; Drosophila.	724	QIQLEKESIEQALAKQNELEDFQKQSEV	HLQEIKAQNTQKDFELVESGESLKLQ	783		
OX	NCBI_TaxID=7227;						
RP	SEQUENCE FROM N.A.	185	-----	ERAVAA-----	ERQRYWEKEBOTLQFKSKMACQLYREK	VNALQAOVCE	231
RC	STRAIN=BERKELEY;						
RX	MEDLINE=20196006; PubMed=10731132;	784	QLEQKTLGHEKLAQALBELKKEKETI	IKERQEQQLQ-SKSA-----	ESESALKVVQVQ	837	
RA	Adams M.D., Ceinliker S.E., Holt R.A., Evans C.A., Gocayne J.D.,						
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	232	LOKEDQAYSARDSAQREISQSLVEKDS	LRQV-----	FELTDQVCELR	TOLRQLOAEP	287
RA	Georgie R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,						
RA	Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,	838	LEQLOQQAASGEESKTVAKLHDEISOL	KSOAEETSELKSTQSNLEAKSKOLEA-	ANG	896	
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,						
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	288	VLKQBAR	294			
RA	Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,						
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	897	SLEEEAK	903			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,						
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,						
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,						
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,						
RA	de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,						
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,						
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferrier A.S., Fleischmann W.,						
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,						
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,						
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,						
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,						
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,						
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,						
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,						
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,						
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,						
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,						
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,						
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,						
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,						
RA	Shue B.C., Sidgen-Klamos I., Simpson M., Skupski M.P., Smith T.,						
RA	Spirer E., Spradling A.C., Stapleton M., Strong R., Sun E.,						
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,						
RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,						
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,						
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,						
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,						
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;						
RT	"The genome sequence of Drosophila melanogaster.";						
RL	Science 287:2185-2195(2000).						
DR	EMBL; AE003655; AAF53604.1; -						
DR	FlyBase; FBgn020503; CLIP-190.						
DR	InterPro; IPR000938; CAP-Gly.						
DR	Pfam; PF01302; CAP_GLY; 2.						
SQ	SEQUENCE 1690 AA; 189063 MW; D6F7916A9C532F16 CRC64;						
	Query Match 13.1%; Score 191.5; DB 5; Length 1690;						
	Best Local Similarity 23.7%; Pred. No. 0.0088;						
	Matches 101; Conservative 69; Mismatches 110; Indels 147; Gaps 19;						
Qy	8 GSLQELN--QEK---GQKEVLLRRCQQLQPH-----	34					
Db	484 GALOEETIAQLQKMTIQOKEVESRIAQLEEQRLRENVKYLNEQIATLQSELVSKDEAL	543					
Qy	35 --LGLAETRAEGLHQ---LEADHSRPMKREVSAPHF-----	EVLRKDEMLSLHY	80				
Db	544 EKPSLSECGIENLRREIALLKENEKQAQAEFTKRLAEKSVFVLRSLSELONLKATS	603					
Qy	81 SNALQEKELAAASRCRSLQELVLLKQELQRANN---VSSCELELQPSLRTASD----	OE	133				
Db	604 DSLESERVNKSDBCEITLQTEVVRMRDEQIRELNQQLDEVTTQLNVQKADSSALDMLRLQK	663					
Qy	134 SGDEELNRLKEENEK-----LRSLTFSLEAKDILQSLQDEARGSRQELVERIHS	LERAV	188				
Db	664 EGTEKSTLLEKTEKELVQTKQEAQNTQKQLEKEISDLKQLAEQKLVREMTNAIN	719					
Qy	189 AAERQ-----REQYWEKEBOTLQFKSKMACQLYREKVN-----	224					
Db	720 NALNQLEKESIEQALAKQNELEDFQKQSEVHLQEIKAQNTQKDFELVESGESLKLQ	779					
Qy	225 -----LQAOVCELOKED-----	QAYSARDSAQREISOL	256				
Db	780 KLOQLEKTLGHEKLAQALBELKKEKETI	IKERQEQQLQSKSAESALKVVQ--	VQ	837			

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 18.125 Seconds
(without alignments)
683.714 Million cell updates/sec

Title: US-09-767-215-2_COPY_568_660
Perfect score: 467
Sequence: 1 QVTMLAFQGDALLEQISVIG.....GLLRVDFGCLSKVNTDQ 93

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				A_Geneseq_101002.*			
1:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1980.DAT.*						
2:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1981.DAT.*						
3:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1982.DAT.*						
4:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1983.DAT.*						
5:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1984.DAT.*						
6:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1985.DAT.*						
7:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1986.DAT.*						
8:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1987.DAT.*						
9:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1988.DAT.*						
10:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1989.DAT.*						
11:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1990.DAT.*						
12:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1991.DAT.*						
13:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1992.DAT.*						
14:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1993.DAT.*						
15:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1994.DAT.*						
16:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1995.DAT.*						
17:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1996.DAT.*						
18:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1997.DAT.*						
19:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1998.DAT.*						
20:	/SIDS2/gcgdata/genseq/genseq-emb1/AA1999.DAT.*						
21:	/SIDS2/gcgdata/genseq/genseq-emb1/AA2000.DAT.*						
22:	/SIDS2/gcgdata/genseq/genseq-emb1/AA2001.DAT.*						
23:	/SIDS2/gcgdata/genseq/genseq-emb1/AA2002.DAT.*						

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	467	100.0	1004	22 AAE07164	Human caspase recr
2	467	100.0	1139	22 AAE07165	Human predicted ca
3	139	29.8	1147	22 AAU01207	Human caspase recr
4	85.5	18.3	1445	22 ABB64619	Drosophila melanog
5	85	18.2	1916	22 ABB62423	Drosophila melanog
6	85	18.2	1916	22 ABB66089	Drosophila melanog
7	80	17.1	548	22 AAU93968	Human stomach canc
8	80	17.1	548	22 AAB33753	Human protein sequ
9	80	17.1	928	23 AAE21718	Human PKIN-13 prot
10	80	17.1	1037	22 ABB22366	Novel human diagno

11	79.5	17.0	1651	23 ABB66725	Human novel polype
12	79.5	17.0	1675	21 AAB42658	Human ORFX ORF2422
13	74.5	16.0	538	21 AAB21149	Human betal-syntro
14	73.5	15.7	540	21 AAB21150	Human beta2-syntro
15	73	15.6	304	22 AAB93296	Human polyptide,
16	73	15.6	763	20 AAY04741	PDZ domain-contain
17	73	15.6	856	21 AAB01383	Neuron-associated
18	73	15.6	1005	20 AAY04731	Mature protein con
19	73	15.6	1373	20 AAY04730	Protein containing
20	73	15.6	1551	22 ABB64459	Drosophila melanog
21	73	15.6	2000	20 AAY04732	Protein containing
22	73	15.6	2037	21 AAY53753	Amino acid sequenc
23	73	15.6	2070	20 AAY04733	Protein containing
24	72.5	15.5	974	22 ABB62642	Drosophila melanog
25	71.5	15.3	229	21 AAG05041	Arabidopsis thalia
26	71.5	15.3	280	21 AAG05040	Arabidopsis thalia
27	71.5	15.3	327	21 AAG05039	Arabidopsis thalia
28	71.5	15.3	503	23 ABB93423	Herbicidally activ
29	71.5	15.3	871	22 ABB67855	Drosophila melanog
30	71.5	15.3	871	22 ABB67265	Drosophila melanog
31	71.5	15.3	1535	23 AAU87934	Human protein cont
32	71.5	15.3	2466	16 AAR71498	Human protein tyro
33	71.5	15.3	2466	19 AAW75999	Intracellular prot
34	71.5	15.3	2466	21 AAY90272	Human PPLI phosph
35	71.5	15.3	2485	21 AAB19343	Amino acid sequenc
36	71	15.2	674	20 AAB92953	Fly transducisome
37	71	15.2	674	22 ABB59957	Drosophila melanog
38	71	15.2	1881	20 AAY24025	Amino acid sequenc
39	70	15.0	96	23 AAB79943	Human PDZPL domain
40	70	15.0	158	22 AAU17383	Novel signal trans
41	70	15.0	272	20 AAY04739	PDZ domain-contain
42	70	15.0	319	20 AAY04740	PDZ domain-contain
43	70	15.0	658	22 ABB09545	Novel human diagno
44	70	15.0	674	19 AAW72748	Human P-dlg protei
45	70	15.0	1111	23 AAU87918	Human PDZPL protei

ALIGNMENTS

RESULT 1
AAE07164
ID AAE07164 standard; Protein; 1004 AA.
XX AAE07164;
AC AAE07164;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human caspase recruitment domain-14 (CARD-14).
XX
KW Human: caspase recruitment domain-14; CARD-14; chromosome 17;
KW nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy;
KW autoimmune disorder; systemic lupus erythematosus; neurological disorder;
KW Alzheimer's disease; Parkinson's disease; inflammatory disorder; anaemia;
KW haematological disorder; myelodysplastic syndrome; myocardial infarction;
KW stroke; immune disorder; Crohn's disease; allergic rhinitis; infection;
KW cell signalling disorder; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiviral; antibacterial.
XX Homo sapiens.
OS

Key Location/Qualifiers
Modified-site 6..9
/note= "CAMP- and cGMP-dependent protein kinase phosphorylation site"
Domain 10..116
/label= CARD_domain
Modified-site 12..15
/note= "Casein kinase II phosphorylation site"
Modified-site 18..21
/note= "Casein kinase II phosphorylation site"
Modified-site 25..27
/note= "Protein kinase C phosphorylation site"

FT	Modified-site	60..62	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	602..605	/note= "N-glycosylation site"
FT	Modified-site	91..93	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	634..637	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	114..117	/note= "N-glycosylation site"	FT	Modified-site	653..655	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	117..122	/note= "N-myristoylation site"	FT	Modified-site	674..677	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	121..123	/note= "Protein kinase C phosphorylation site"	FT	Domain	676..745	/label= SH3_domain
FT	Domain	126..420	/label= Coiled_Coil_domain	FT	Modified-site	714..719	/note= "N-myristoylation site"
FT	Modified-site	130..135	/note= "N-myristoylation site"	FT	Modified-site	725..727	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	134..137	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	725..728	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	161..166	/note= "N-myristoylation site"	FT	Modified-site	733..738	/note= "N-myristoylation site"
FT	Modified-site	165..168	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	737..740	/note= "N-glycosylation site"
FT	Modified-site	220..227	/note= "Tyrosine kinase phosphorylation site"	FT	Modified-site	759..761	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	221..224	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	760..763	/note= "CAMP- and cGMP-dependent protein kinase phosphorylation site"
FT	Domain	239..325	/label= k-Box_domain	FT	Peptide	785..793	/note= "Peroxisomal targeting signal"
FT	Modified-site	240..243	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	796..799	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	250..252	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	800..805	/note= "N-myristoylation site"
FT	Modified-site	253..256	/note= "Casein kinase II phosphorylation site"	FT	Domain	826..1004	/label= Guanylate_kinase_domain
FT	Modified-site	259..262	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	842..844	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	280..283	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	860..863	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	290..293	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	868..870	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	297..300	/note= "Casein kinase II phosphorylation site"	FT	Region	870..872	/note= "RGD cell attachment sequence"
FT	Modified-site	307..309	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	893..896	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	307..310	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	926..929	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	359..365	/note= "Tyrosine kinase phosphorylation site"	FT	Peptide	941..949	/note= "Peroxisomal targeting signal"
FT	Modified-site	366..368	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	944..947	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	366..369	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	976..979	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	378..381	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	980..985	/note= "N-myristoylation site"
FT	Modified-site	384..396	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	1002..1004	/note= "Protein kinase C phosphorylation site"
FT	Region	385..406	/note= "Leucine zipper pattern"	FT	FT	WO200159065-A2.	
FT	Modified-site	449..452	/note= "Casein kinase II phosphorylation site"	FT	XX	16-AUG-2001.	
FT	Modified-site	463..466	/note= "Casein kinase II phosphorylation site"	FT	XX	22-JAN-2001; 2001WO-US02087.	
FT	Modified-site	463..465	/note= "Protein kinase C phosphorylation site"	FT	XX	09-FEB-2000; 2000US-0181159.	
FT	Modified-site	470..472	/note= "Protein kinase C phosphorylation site"	FT	XX	(MILL-) MILLENNIUM PHARM INC.	
FT	Modified-site	501..504	/note= "Casein kinase II phosphorylation site"	FT	XX	Bertin J;	
FT	Modified-site	511..516	/note= "N-myristoylation site"	FT	XX	WPI; 2001-497073/54.	
FT	Domain	568..660	/label= PD2_domain	FT	XX	N-PSDB; AAD13447.	
FT	Modified-site	587..592	/note= "N-myristoylation site"	FT	XX	An isolated caspase recruitment domain polypeptide useful for	
FT	Modified-site	589..592		FT			

PT regulating growth and cell death and useful for the treatment of cancer
PT
XX
PS
XX
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

Sequence 1004 AA;
Query Match 100.0%; Score 467; DB 22; Length 1004;
Best Local Similarity 100.0%; Pred. No. 2.2e-51;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QVTMLAFQGDALLEQISVIGGNLTGIFHRVTPGSAADQMALRPGTQIVMVDYEAASEPLF 60
Db 568 QVTMLAFQGDALLEQISVIGGNLTGIFHRVTPGSAADQMALRPGTQIVMVDYEAASEPLF 627

QY 61 KAVLEDTTLEEAVGLLRVDGFCCLSVKVTNDG 93
Db 628 KAVLEDTTLEEAVGLLRVDGFCCLSVKVTNDG 660

RESULT 2
AAE07165
ID AAE07165 standard; Protein; 1139 AA.
XX
AC AAE07165;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human predicted caspase recruitment domain-14 (CARD-14).
XX
KW Human; caspase recruitment domain-14; CARD-14; chromosome 17;
KW nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy;
KW autoimmune disorder; systemic lupus erythematosus; neurological disorder;
KW Alzheimer's disease; Parkinson's disease; inflammatory disorder; anaemia;
KW haematological disorder; myelodysplastic syndrome; myocardial infarction;
KW stroke; immune disorder; Crohn's disease; allergic rhinitis; infection;
KW cell signalling disorder; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiviral; antibacterial.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 700 /note= "Encoded by TGG"
FT
FT
XX
PN WO200159065-A2.
XX
XX 16-AUG-2001.
XX
XX 22-JAN-2001; 2001WO-US02087.
XX
XX 09-FEB-2000; 2000US-0181159.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Bertin J;
XX
XX WPI; 2001-497073/54.
DR N-PSDB; AAD13448.
XX

PT An isolated caspase recruitment domain polypeptide useful for
PT regulating growth and cell death and useful for the treatment of cancer
XX
XX
PS Disclosure; Fig 2A-2C; 109pp; English.
XX
CC The present sequence is predicted human caspase recruitment domain-14
CC (CARD-14). The CARD-14 gene is located on chromosome 17. The CARD-14 is
CC used for the detection of modulators that modulates the ability of
CC CARD-14 to bind to Bcl-10 and stimulates phosphorylation of Bcl-10 or
CC activation of nuclear factor-kappa B (NF-kB). The CARD-14 is useful for
CC regulating growth and cell death and useful for the treatment of cancer.
CC It is also useful for the treatment of autoimmune disorders (e.g.,
CC systemic lupus erythematosus), neurological disorders, haematological disorders
CC and Parkinson's disease, inflammatory disorders, e.g., Alzheimer's
CC (e.g., anaemia, myelodysplastic syndromes), myocardial infarctions,
CC strokes, immune disorders (e.g., Crohn's disease, allergic rhinitis),
CC cell signalling disorders and certain viral and bacterial infections.
XX
SQ Sequence 1139 AA;
Query Match 100.0%; Score 467; DB 22; Length 1139;
Best Local Similarity 100.0%; Pred. No. 2.6e-51;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QVTMLAFQGDALLEQISVIGGNLTGIFHRVTPGSAADQMALRPGTQIVMVDYEAASEPLF 60
Db 618 QVTMLAFQGDALLEQISVIGGNLTGIFHRVTPGSAADQMALRPGTQIVMVDYEAASEPLF 677

QY 61 KAVLEDTTLEEAVGLLRVDGFCCLSVKVTNDG 93
Db 678 KAVLEDTTLEEAVGLLRVDGFCCLSVKVTNDG 710

RESULT 3
AAU01207
ID AAU01207 standard; Protein; 1147 AA.
XX
AC AAU01207;
XX
DT 12-SEP-2001 (first entry)
XX
DE Human caspase recruitment domain, CARD-11 polypeptide.
XX
KW Human; caspase recruitment domain; CARD-11; Bcl-10; NF-kappaB;
KW apoptosis; hyperproliferative disorder; autoimmune; neurological;
KW inflammatory disorder; viral infection; stress-related response.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 6..112 /note= "CARD domain"
FT Modified-site 7..9 /note= "Protein kinase C phosphorylation site"
FT Modified-site 7..10 /note= "Protein kinase II phosphorylation site"
FT Modified-site 100..102 /note= "Protein kinase C phosphorylation site"
FT Modified-site 100..103 /note= "Casein kinase II phosphorylation site"
FT Modified-site 105..107 /note= "Protein kinase C phosphorylation site"
FT Modified-site 106..109 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Domain 130..431 /note= "Coiled coil domain"
FT Modified-site 162..165 /note= "Casein kinase II phosphorylation site"
FT Modified-site 168..171 /note= "Casein kinase II phosphorylation site"
FT Modified-site 175..183

FT /note= "Tyrosine kinase phosphorylation site"
FT 182..185
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 189..195
FT Modified-site
FT /note= "Tyrosine kinase phosphorylation site"
FT 241..244
FT Modified-site
FT /note= "N-glycosylation site"
FT 243..245
FT Modified-site
FT /note= "Protein kinase C phosphorylation site"
FT 282..285
FT Modified-site
FT /note= "Amidation site"
FT 286..289
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 290..292
FT Modified-site
FT /note= "Protein kinase C phosphorylation site"
FT 378..381
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 429..432
FT Modified-site
FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT 459..461
FT Modified-site
FT /note= "Protein kinase C phosphorylation site"
FT 471..474
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 472..475
FT Modified-site
FT /note= "N-glycosylation site"
FT 476..479
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 508..510
FT Modified-site
FT /note= "Protein kinase C phosphorylation site"
FT 510..513
FT Modified-site
FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT 558..560
FT Modified-site
FT /note= "Protein kinase C phosphorylation site"
FT 578..581
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 584..587
FT Modified-site
FT /note= "N-glycosylation site"
FT 587..592
FT Modified-site
FT /note= "N-myristoylation site"
FT 634..637
FT Modified-site
FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT 635..748
FT Domain
FT /note= "PDZ domain"
FT 635..1147
FT Domain
FT /note= "MAGUK domain"
FT 638..641
FT Modified-site
FT /note= "Glycosaminoglycan attachment site"
FT 678..683
FT Modified-site
FT /note= "N-myristoylation site"
FT 687..689
FT Modified-site
FT /note= "Protein kinase C phosphorylation site"
FT 692..695
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 698..703
FT Modified-site
FT /note= "N-myristoylation site"
FT 710..715
FT Modified-site
FT /note= "N-myristoylation site"
FT 725..728
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 761..766
FT Modified-site
FT /note= "N-myristoylation site"
FT 764..767
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 766..834
FT Domain
FT /note= "SH3 domain"
FT 776..779
FT Modified-site
FT /note= "N-glycosylation site"
FT 779..782
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 787..789

FT /note= "Protein kinase C phosphorylation site"
FT 816..819
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 823..828
FT Modified-site
FT /note= "N-myristoylation site"
FT 847..850
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 853..858
FT Modified-site
FT /note= "N-myristoylation site"
FT 857..859
FT Modified-site
FT /note= "Protein kinase C phosphorylation site"
FT 872..875
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 882..1147
FT Domain
FT /note= "Guanylate kinase (GUK) domain"
FT 897..900
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 917..922
FT Modified-site
FT /note= "N-myristoylation site"
FT 926..929
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 935..937
FT Modified-site
FT /note= "Protein kinase C phosphorylation site"
FT 1003..1006
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 1010..1018
FT Modified-site
FT /note= "Tyrosine kinase phosphorylation site"
FT 1050..1055
FT Modified-site
FT /note= "N-myristoylation site"
FT 1088..1091
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT 1120..1123
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT
XX WO200140468-A2.
XX
XX 07-JUN-2001.
XX
XX 01-DEC-2000; 2000WO-US32716.
XX
XX 03-DEC-1999; 99US-0168780.
PR 18-FEB-2000; 2000US-0507533.
PR 25-FEB-2000; 2000US-0513904.
PR 10-OCT-2000; 2000US-0685791.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Bertin J;
XX
XX WPI; 2001-367809/38.
DR N-PSDB; AAS05389.
XX
XX Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10, CARD-11, useful as targets for therapy, as immunogens, and in screening and detection assays -
XX
XX Claim 9; Fig 14A-14C; 145pp; English.
XX
XX The present sequence represents novel human caspase recruitment domain, CARD-11. The polynucleotide encoding this sequence was isolated from a human T-cell cDNA library. Also described are novel human sequences for CARD-9 and CARD-10 (AAU01205, AAU01206) and rat CARD-9 (AAU01204). CARD-9, CARD-10 and CARD-11 interact with Bcl-10 which is thought to activate nuclear factor (NF)-kappaB and apoptosis. The sequences of the invention can be used for treating a disorder associated with abnormal levels of apoptosis by modulating the expression or activity of CARD-9, CARD-10, or CARD-11. They can be used for the treatment of hyperproliferative disorders (e.g. cancer), autoimmune disorders (e.g. systemic lupus erythematosus), neurological disorders (e.g. Alzheimer's disease), inflammatory disorders (e.g. Crohn's disease), and viral infection (e.g. HIV). The CARD polypeptide, polynucleotide and an antibody which selectively binds to CARD can be used in screening and detection assays (e.g. chromosomal


```

QY      76 LRRV 79
      : ::
Db      1564 ISKL 1567

RESULT 6
ABB66089
ID      ABB66089 standard; Protein; 1916 AA.
XX      AC
XX      ABB66089;
XX      DT
XX      26-MAR-2002 (first entry)
XX      DE
XX      Drosophila melanogaster polypeptide SEQ ID NO 25059.
XX      KW
XX      Drosophila; developmental biology; cell signalling; insecticide;
XX      pharmaceutical.
XX      OS
XX      Drosophila melanogaster.
XX      WO200171042-A2.
XX      PN
XX      XX
XX      PD
XX      27-SEP-2001.
XX      PF
XX      23-MAR-2001; 2001WO-US09231.
XX      PR
XX      23-MAR-2000; 2000US-191637P.
XX      PR
XX      11-JUL-2000; 2000US-0614150.
XX      PA
XX      (PEKE ) PE CORP NY.
XX      PI
XX      Venter JC, Adams M, Li PWD, Myers EW;
XX      WPI; 2001-656860/75.
XX      DR
XX      N-PSDB; ABL10192.
XX      PT
XX      New isolated nucleic acid detection reagent for detecting 1000 or more
XX      genes from Drosophila and for elucidating cell signalling and cell-cell
XX      interactions -
XX      PS
XX      Disclosure; SEQ ID NO 25059; 21pp + Sequence Listing; English.
XX      CC
XX      The invention relates to an isolated nucleic acid detection reagent
XX      capable of detecting 1000 or more genes from Drosophila. The invention is
XX      useful in developmental biology and in elucidating cell signalling and
XX      cell-cell interactions in higher eukaryotes for the development of
XX      insecticides, therapeutics and pharmaceutical drugs. The invention
XX      discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA
XX      sequences (ABU01840-ABU16175) and the encoded proteins
XX      (ABB57737-ABB72072).
XX      CC
XX      The sequence data for this patent did not form part of the printed
XX      CC specification, but was obtained in electronic format directly from WIPO
XX      CC at ftp.wipo.int/pub/published_pct_sequences.
XX      SQ
XX      Sequence 1916 AA;

      Query Match      18.2%; Score 85; DB 22; Length 1916;
      Best Local Similarity 31.2%; Pred. No. 0.28;
      Matches 20; Conservative 12; Mismatches 24; Indels 8; Gaps 2;

QY      16 ISVIGNLTGIFTHRVTPGSAADQMALRPGTQIVMVYDEASEPLFKAVLEDTTLEAVGL 75
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1512 IKLFGNKVGIIHVHDVAVGSPDHAGIRKGDQI--LEYNQVD-----LSGYTAEQAANE 1563
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      76 LRRV 79
      : ::
Db      1564 ISKL 1567

RESULT 7
AAM93968
ID      AAM93968 standard; Protein; 548 AA.
XX

```

XX OS Homo sapiens.
XX PN EPI074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX PR Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX PT
XX PS Claim 8; SEQ ID 13422; 2537pp + CD ROM; English.
XX CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX SQ Sequence 548 AA;
Query Match 17.1%; Score 80; DB 22; Length 548;
Best Local Similarity 34.4%; Pred. No. 0.22;
Matches 21; Conservative 9; Mismatches 23; Indels 8; Gaps 1;
QY 16 ISVTGGNLTGFIHVRPGSAADQMALRPGTQIVWVDYEASEPLFKAVLEDTTLEAVGL 75
Db 23 LRLAGNDVGI FVSGVQAGSPAGGQIGQDQILQVN-----DVPFQNLTRREAVQF 74
Qy 76 L 76
Db 75 L 75
RESULT 9
AAE21718
ID AAE21718 standard; Protein: 928 AA.
XX AC AAE21718;
XX

DT 16-JUL-2002 (first entry)
XX DE Human PKIN-13 protein.
XX KW Human; kinase; enzyme; PKIN-13 protein; immune system disorder; anaemia;
KW acquired immune deficiency syndrome; thymic hypoplasia; Crohn's disease;
KW asthma; neurological disorder; epilepsy; Charcot-Marie-Tooth disease;
KW AIDS; seizures; cell proliferative disorder; cancer; adenocarcinoma;
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; developmental disorder;
KW Down's syndrome; gene therapy; protein therapy; cytostatic.
XX OS Homo sapiens.
XX FH Key
FT Domain Location/Qualifiers
FT 10..99 /note= "GLGF domain"
FT 20..101 /note= "PDZ domain"
FT 204..280 /note= "PDZ domain"
FT 388..469 /note= "PDZ domain"
FT 391..471 /note= "GLGF domain"
FT 429..439 /note= "PDZ domain"
FT 733..754 /note= "PDZ domain"
FT /note= "Leucine zipper domain"
XX PN WO200218557-A2.
XX PD 07-MAR-2002.
XX PF 31-AUG-2001; 2001WO-US27219.
XX PR 31-AUG-2000; 2000US-229873P.
XX PR 08-SEP-2000; 2000US-231357P.
XX PR 14-SEP-2000; 2000US-232654P.
XX PR 22-SEP-2000; 2000US-234902P.
XX PR 29-SEP-2000; 2000US-236499P.
XX PR 06-OCT-2000; 2000US-238389P.
XX PR 13-OCT-2000; 2000US-240542P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Bandman O, Nguyen DB, Walia NK, Hafalia AJA, Yao MG, Gandhi AR;
PI Gururajan R, Ding L, Patterson C, Yue H, Baughn MR, Tribouley CM;
PI Thornton M, Elliott VS, Lu Y, Ison CH, Au-Young J, Tang YT;
PI Azimzai Y, Burrill JD, Marcus GA, Zingler KA, Lu DAM, Lal PG;
PI Rankumar J, Warren BA, Kearney L, Policky JL, Thangavelu K;
PI Burford N;
XX WPI: 2002-329769/36.
XX N-PSDB: AAD34310.
XX PT New human kinases, useful for diagnosing, treating or preventing immune
PT system disorders (e.g. Crohn's disease), neurological disorders (e.g.
PT epilepsy), or cell proliferative disorders (e.g. cancers such as
PT leukemia or lymphoma) -
XX PS Claim 68; Page 174-176; 218pp; English.
XX CC The present invention relates to human kinases (PKIN) and polynucleotides
XX encoding such proteins. PKIN sequences of the invention are useful for
XX diagnosing, treating or preventing disorders associated with aberrant
XX expression of PKIN, particularly immune system disorders (e.g. acquired
XX immune deficiency syndrome (AIDS), thymic hypoplasia, Crohn's disease,
XX anaemia, asthma), neurological disorders (e.g. epilepsy, Charcot-Marie-
XX Tooth disease or seizures), cell proliferative disorders (e.g. cancers
XX such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma),
XX and developmental disorders (e.g. Down's syndrome). They are also used
XX in gene therapy and protein therapy. The present sequence is human
XX PKIN-13 protein.

XX
SQ Sequence 928 AA;

Query Match 17.1%; Score 80; DB 23; Length 928;
Best Local Similarity 34.4%; Pred. No. 0.45;
Matches 21; Conservative 9; Mismatches 23; Indels 8; Gaps 1;

QY 16 ISVIGGNLTGIFHRVTPGSAADOMALRPGTOIVWVDYEAASEPLFKAVLEDTTLEEAAGL 75
: : ||| ||| : || || : : || : : : |||
Db 403 LRLAGGNDVGIFVSGVQAGSPADGGIOGDILOVN-----DVPFONLTREEAVQF 454

QY 76 L 76
Db 455 L 455

RESULT 10
ABG22366
ID ABG22366 standard; Protein; 1037 AA.
XX
AC ABG22366;
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #22357.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS86553.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 52725; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 1037 AA;

Query Match 17.1%; Score 80; DB 22; Length 1037;
Best Local Similarity 34.4%; Pred. No. 0.52;
Matches 21; Conservative 9; Mismatches 23; Indels 8; Gaps 1;

QY 16 ISVIGGNLTGIFHRVTPGSAADOMALRPGTOIVWVDYEAASEPLFKAVLEDTTLEEAAGL 75
: : ||| ||| : || || : : || : : : |||

Db 506 LRLAGGNDVGIFVSGVQAGSPADGGIOGDILOVN-----DVPFONLTREEAVQF 557

QY 76 L 76

Db 558 L 558

RESULT 11

ABG66725

ID ABG66725 standard; Protein; 1651 AA.

XX
AC ABG66725;

DT 30-AUG-2002 (first entry)

XX
DE Human novel polypeptide #60.

XX
KW Human: inflammatory condition; shock; sepsis; immune response;

KW cancer; wound healing; central nervous system disease; haematopoiesis;

KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;

KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;

KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;

KW bone degenerative disorder; periodontal disease; reperfusion injury;

KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;

KW allergic condition; thrombolysis; thrombosis; coagulation disorder;

KW fungal infection.

XX
OS Homo sapiens.

XX
PN WO200244340-A2.

XX
PD 06-JUN-2002.

XX
PF 30-NOV-2001; 2001WO-US47004.

XX
PR 30-NOV-2000; 2000US-0028952.

XX
PA (HYSE-) HYSEQ INC.

XX
PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;

XX
PI Yamazaki V, Ujwal ML, Drmanac RT;

XX
DR WPI: 2002-508509/54.

XX
DR N-PSDB; ABK94949.

XX
PT Novel nucleic acids and polypeptides for diagnosis, treatment of

PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell

PT disorders, cancer and promoting wound healing -

XX
PS Claim 10; Page 629-632; 672pp; English.

XX
CC The invention relates to human novel polynucleotides and associated

CC polypeptides. The polynucleotides and polypeptides are useful for

CC treating inflammatory conditions such as arthritis, nephritis, Crohn's

CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses

CC and cancer and for promoting wound healing. The sequences are used to

CC induce the proliferation of neural cells and regeneration of nerve and

CC brain tissue, and are useful for the treatment of central and peripheral

CC nervous system diseases and neuropathies, such as Alzheimer's disease,

CC Parkinson's disease, Huntington's disease and amyotrophic lateral

CC sclerosis. The sequences are involved in chemotactic or chemokinetic

CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid

CC cell disorders and platelet disorders such as thrombocytopenia,
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC disease. The sequences of the invention are also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC and coagulation disorders. Sequences ABG66666-ABG66758 represent human
CC novel polypeptides of the invention.

XX Sequence 1651 AA;
SQ
Query Match 17.0%; Score 79.5; DB 23; Length 1651;
Best Local Similarity 34.6%; Pred. No. 1.2;
Matches 27; Conservative 12; Mismatches 28; Indels 11; Gaps 3;
QY 15 QISVIGGNLT--GIFHRTVTPGSAADQMALRPGTQIWMVDYEAASEPLFKAVLEDTTLEA 72
Db 593 QFSLGGSEKGFIFVEGVEPGSKAADSGLKRGDQIMEVNGQN-----FENITFMKA 644
QY 73 VGLLRVDGFCCLSVKVN 90
Db 645 VEILRN-NTHLALTVKTN 661

RESULT 12
AAB42658
ID AAB42658 standard; Protein: 1675 AA.
XX
AC AAB42658;
DT 08-FEB-2001 (first entry)
DE Human ORF2422 polypeptide sequence SEQ ID NO:4844.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antineoplastic; antithyroid;
KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

OS Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
PA
PA Shimkets RA, Leach M;
PI
XX
XX WPI; 2000-602362/57.
DR
DR N-PSDB; AAC76867.
XX

PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 4028-4032; 5507pp; English.
CC AAB74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 1675 AA;
SQ
Query Match 17.0%; Score 79.5; DB 21; Length 1675;
Best Local Similarity 34.6%; Pred. No. 1.2;
Matches 27; Conservative 12; Mismatches 28; Indels 11; Gaps 3;
QY 15 QISVIGGNLT--GIFHRTVTPGSAADQMALRPGTQIWMVDYEAASEPLFKAVLEDTTLEA 72
Db 616 QFSLGGSEKGFIFVEGVEPGSKAADSGLKRGDQIMEVNGQN-----FENITFMKA 667
QY 73 VGLLRVDGFCCLSVKVN 90
Db 668 VEILRN-NTHLALTVKTN 684

RESULT 13
AAB21149
ID AAB21149 standard; peptide; 538 AA.
XX
AC AAB21149;
DT 21-FEB-2001 (first entry)
XX
XX Human betal-syntrophin.
DE
XX
XX Human; betal-syntrophin; SNT BI; PDZ domain; GLGF loop;
KW DHF domain; muscular dystrophy.

OS Homo sapiens.
XX
XX WO200048002-A1.
XX
XX 17-AUG-2000.
XX
XX 09-FEB-2000; 2000WO-GB00374.
XX
XX 09-FEB-1999; 99GB-0002696.
XX
XX (MEDI-) MEDICAL RES COUNCIL.
PA
XX
XX Spillantini MG, Goedert M, Hasegawa M, Buee-Scherrer V, Thomas G;
PI Cohen P, Cuenda A;
XX
XX WPI; 2000-558226/51.
XX

Search completed: January 22, 2003, 08:52:19
Job time : 21.125 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:52:30 ; Search time 6.375 Seconds
(without alignments)
429.229 Million cell updates/sec

Title: us-09-767-215-2_COPY_568_660

Perfect score: 467

Sequence: 1 QVTMLAFQDALLQISVIG.....GLLRVDFGFCCLSVKVNVDG 93

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PTOUS_COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73	15.6	2037	US-09-306-998-3	Sequence 3, Appli
2	72	15.4	2465	US-08-596-291-3	Sequence 3, Appli
3	72	15.4	2465	US-09-100-804-3	Sequence 3, Appli
4	71.5	15.3	2466	US-09-080-855-12	Sequence 12, Appli
5	71.5	15.3	2466	PCT-US94-09943-2	Sequence 2, Appli
6	71.5	15.3	2485	US-09-290-640-46	Sequence 46, Appli
7	71	15.2	1881	US-09-233-086-3	Sequence 3, Appli
8	69.5	14.9	77	US-09-100-804-16	Sequence 16, Appli
9	66.5	14.2	173	US-08-923-454A-2	Sequence 2, Appli
10	66.5	14.2	316	US-08-739-485-5	Sequence 5, Appli
11	66.5	14.2	458	US-08-923-454A-8	Sequence 8, Appli
12	66.5	14.2	458	US-08-923-454A-25	Sequence 25, Appli
13	66.5	14.2	458	US-08-923-454A-31	Sequence 31, Appli
14	66.5	14.2	458	US-09-008-271A-11	Sequence 11, Appli
15	66	14.1	539	US-08-759-581B-20	Sequence 20, Appli
16	66	14.1	539	US-09-304-711-20	Sequence 20, Appli
17	66	14.1	539	US-09-173-281-20	Sequence 20, Appli
18	63.5	13.6	297	US-09-151-611-3	Sequence 3, Appli
19	63.5	13.6	297	US-09-370-102-3	Sequence 3, Appli
20	63.5	13.6	505	US-08-123-161A-14	Sequence 14, Appli
21	63.5	13.6	505	US-08-483-278-14	Sequence 14, Appli
22	61	13.1	858	US-08-946-026-3	Sequence 3, Appli
23	60.5	13.0	1642	US-08-447-411-45	Sequence 45, Appli
24	60.5	13.0	1642	US-08-662-227-2	Sequence 2, Appli
25	60.5	13.0	1642	US-09-017-947-2	Sequence 2, Appli
26	60.5	13.0	1648	US-08-662-227-35	Sequence 35, Appli
27	60.5	13.0	1648	US-09-017-947-35	Sequence 35, Appli

28	60	12.8	724	4	US-09-562-737-26	Sequence 26, Appli
29	59.5	12.7	374	4	US-09-091-405-2	Sequence 2, Appli
30	59.5	12.7	631	4	US-09-147-119-7	Sequence 7, Appli
31	59	12.6	71	3	US-09-100-804-31	Sequence 31, Appli
32	59	12.6	80	3	US-08-545-860D-52	Sequence 52, Appli
33	59	12.6	80	5	PCT-US94-04496-52	Sequence 5, Appli
34	59	12.6	394	2	US-08-530-290-23	Sequence 23, Appli
35	59	12.6	395	2	US-08-990-379-5	Sequence 5, Appli
36	58.5	12.5	560	4	US-09-385-742B-6	Sequence 6, Appli
37	58.5	12.5	1333	1	US-08-447-411-76	Sequence 76, Appli
38	58.5	12.5	1333	2	US-08-662-227-34	Sequence 34, Appli
39	58.5	12.5	1333	4	US-09-017-947-34	Sequence 34, Appli
40	58	12.4	724	4	US-09-562-737-23	Sequence 23, Appli
41	58	12.4	1548	4	US-09-376-330-15	Sequence 15, Appli
42	57.5	12.3	582	4	US-09-413-814-83	Sequence 83, Appli
43	57	12.2	463	1	US-08-278-091-2	Sequence 2, Appli
44	57	12.2	463	1	US-08-483-859-2	Sequence 2, Appli
45	57	12.2	463	1	US-08-472-173-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-306-998-3
; Sequence 3, Application US/09306998
; Patent No. 6291173
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: MMS2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMS2
; CURRENT APPLICATION NUMBER: US/09/306,998
; CURRENT FILING DATE: 1999-05-07
; EARLIER APPLICATION NUMBER: US 60/084,740
; EARLIER FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-306-998-3

Query Match	15.6%	Score 73:	DB 4;	Length 2037;
Best Local Similarity	31.2%	Pred. No. 1.4;		
Matches	24;	Conservative	14;	Mismatches 25; Indels 14; Gaps 3;
QY	17	SVIGG-----NLTGIFIRHV-TPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLE	70	
Db	1967	SIVGGYSPHGDLPIYVKTVFKAASEDGLKRGDIIAVNGQS-----LEGVTHE	2018	
QY	71	EAVGLLRVDGFCCLSV	87	
Db	2019	EAVAILKRTKGTVTLMV	2035	
RESULT 2				
US-08-596-291-3				
; Sequence 3, Application US/08596291				
; Patent No. 5821075				
; GENERAL INFORMATION:				
; APPLICANT: GONEZ, LEONEL JORGE				
; APPLICANT: SARAS, JAN				
; APPLICANT: CLAESON-WELSH, LENA				
; APPLICANT: HELDIN, CARL-HENRIK				
; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL				
; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN				
; TITLE OF INVENTION: TYROSINE PHOSPHATASES				
; NUMBER OF SEQUENCES: 4				
; CORRESPONDENCE ADDRESS:				
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.				
; STREET: 600 ATLANTIC AVENUE				

```

CLASSIFICATION:
PRIORITY APPLICATION DATA: US 08/596,291
APPLICATION NUMBER: 09-AUG-1996
FILING DATE: 09-AUG-1996
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09943
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: LO461/7003
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2465 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-100-804-3

Query Match 15.4%; Score 72; DB 3; Length 2465;
Best Local Similarity 28.6%; Pred. No. 2.5;
Matches 24; Conservative 15; Mismatches 30; Indels 14; Gaps 4;

QY 1 QVTMLAFQGDALLE-QISVIGNL-----TGFIHRVTGSAAD-QMALRPGTQIVMVD-- 52
Ddb 1071 EITLVNLRKDAKYGVLGQIIGGKMETDLGFISSVAPGPGADFHGCLKPGDRLISVNSV 1130
QY 53 -----YEASEPLFKAVLEDTTL 69
Ddb 1131 SLEGVSHHAAIEILQNAPEDVTL 1153

RESULT 4
US-09-080-855-12
Sequence 12, Application 05/09080855A
Patent No. 6083721
GENERAL INFORMATION:
APPLICANT: Saras, Jan
APPLICANT: Franz, Petra
APPLICANT: Aspenstrm, Pontus
APPLICANT: Hellman, Ulf
APPLICANT: Gonez, Leonel Jorge
APPLICANT: Heldin, Carl-Henrik
TITLE OF INVENTION: PARG, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTPL1
FILE REFERENCE: LO461/7030
CURRENT APPLICATION NUMBER: US/09/080,855A
CURRENT FILING DATE: 1998-05-18
EARLIER APPLICATION NUMBER: 08/805,583
EARLIER FILING DATE: 1997-02-25
NUMBER OF SEQ ID NOS: 39
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 2466
TYPE: PRT
ORGANISM: Homo sapiens
US-09-080-855-12

Query Match 15.3%; Score 71.5; DB 3; Length 2466;
Best Local Similarity 28.6%; Pred. No. 2.9;
Matches 24; Conservative 15; Mismatches 30; Indels 15; Gaps 4;

QY 1 QVTMLAFQGDALLE-QISVIGNL-----TGFIHRVTGSAAD-QMALRPGTQIVMVD- 52
Ddb 1071 EITLVNLRKDAKYGVLGQIIGGKMETDLGFISSVAPGPGADFHGCLKPGDRLISVNS 1130
QY 53 -----YEASEPLFKAVLEDTTL 69

```



```
; REFERENCE/DOCKET NUMBER: PF-0142 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
US-08-739-485-5

Query Match 14.2%; Score 66.5; DB 2; Length 316;
Best Local Similarity 32.5%; Pred. No. 0.74;
Matches 27; Conservative 12; Mismatches 35; Indels 9; Gaps 3;

Qy 20 GGNLTG-----IFIHRTVPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEA 72
Db 14 GFRLSGGIDFNQPLVITRTTIGSKRAANLCPGDVILAIDGFGTSTWTHADAQRIKAAA 73

Qy 73 VGLLRVD-GFCCL-SVKVNTDG 93
Db 74 HOLCLKIDRGTHLWSPQVSDG 96

RESULT 11
US-08-923-454A-8
; Sequence 8, Application US/08923454A
; Patent No. 6004794
; GENERAL INFORMATION:
; APPLICANT: Creasy, Caretha
; APPLICANT: Livi, George
; APPLICANT: Karran, Eric
; APPLICANT: Clinckenbeard, Helen
; APPLICANT: Browne, Michael
; APPLICANT: Southan, Christopher
; TITLE OF INVENTION: HUMAN SERINE PROTEASE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923.454A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/025436
; FILING DATE: 06-SEPT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50547
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 458 amino acids
; TYPE: amino acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
US-08-923-454A-8

Query Match 14.2%; Score 66.5; DB 3; Length 458;
Best Local Similarity 33.3%; Pred. No. 1.3;
Matches 16; Conservative 11; Mismatches 18; Indels 3; Gaps 1;

Qy 25 GIFHRTVPGSAADQMALRPGTQIVMVD---YEASEPLFKAVLEDTTTL 69
Db 390 GVLIHKVILGSPAHRAGLRPGDVILAIGEOMVQNAEDVVEAVRTQSQL 437

RESULT 12
US-08-923-454A-25
; Sequence 25, Application US/08923454A
; Patent No. 6004794
; GENERAL INFORMATION:
; APPLICANT: Creasy, Caretha
; APPLICANT: Livi, George
; APPLICANT: Karran, Eric
; APPLICANT: Clinckenbeard, Helen
; APPLICANT: Browne, Michael
; APPLICANT: Southan, Christopher
; TITLE OF INVENTION: HUMAN SERINE PROTEASE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923.454A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/025436
; FILING DATE: 06-SEPT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50547
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 458 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
US-08-923-454A-25

Query Match 14.2%; Score 66.5; DB 3; Length 458;
Best Local Similarity 33.3%; Pred. No. 1.3;
```

RESULT 14
US-09-008-271A-11

APPLICANT: WARREN, PATRICK V.
TITLE OF INVENTION: METHODS FOR IDENTIFYING HERBICIDAL AGENTS THAT INHIBIT D1 PR

NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
SOFTWARE: MICROSOFT WORD 2.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/759,581B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 539 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
STRAIN: SPINACH D1 PROTEASE PROTEIN
US-08-759-581B-20

Query Match 14.1%; Score 66; DB 2; Length 539;
Best Local Similarity 24.7%; Pred. No. 1.9;
Matches 18; Conservative 16; Mismatches 25; Indels 14; Gaps 2;
Qy 9 GDALLEQISVIGGNLTGTFIHRVTPGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTT 68
Db 241 GPTAVDQSS-----TGLVVISATPGAPASRAGILPGDVLAIIDASTDKM-----G 286
Qy 69 LEEAVGLLRVDG 81
Db 287 IYEAANILQPGDG 299

Search completed: January 22, 2003, 08:57:37
Job time : 8.375 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 3.75 Seconds
(without alignments)
500.428 Million cell updates/sec

Title: US-09-767-215-2_COPY_568_660
Perfect score: 467
Sequence: 1 QVTMLAFQGDALLEQISVIG.....GLLRVDGFCCLSVKVTDTG 93

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues
Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	467	100.0	1004	10	US-09-767-215-2
2	467	100.0	1138	10	US-09-767-215-5
3	133	28.5	1113	9	US-10-032-159A-14
4	133	28.5	1247	9	US-10-032-159A-8
5	77.5	16.6	80	9	US-09-963-959-11
6	73	15.6	2037	9	US-09-951-402-3
7	73	15.6	2037	10	US-09-951-401-3
8	73	15.6	2037	10	US-09-922-101-3
9	71.5	15.3	2485	10	US-09-802-669-46
10	71	15.2	1881	9	US-09-998-425-3
11	70	15.0	65	10	US-09-911-826A-20
12	70	15.0	158	9	US-09-764-868-948
13	68.5	14.7	77	10	US-09-911-826A-16
14	68.5	14.7	163	9	US-09-764-868-964
15	66.5	14.2	316	9	US-10-126-099-5
16	66.5	14.2	339	10	US-09-925-300-1508
17	64.5	13.8	1499	10	US-09-911-826A-2
18	64	13.7	926	9	US-10-023-437-57
19	63.5	13.6	71	10	US-09-911-826A-17

20	63.5	13.6	297	10	US-09-909-005-3
21	62	13.3	98	10	US-09-739-907-60
22	62	13.3	113	10	US-09-739-907-119
23	62	13.3	1701	9	US-09-963-959-2
24	61	13.1	167	9	US-09-764-868-953
25	61	13.1	314	10	US-09-771-730-22
26	61	13.1	314	10	US-09-771-730-24
27	61	13.1	314	10	US-09-771-730-26
28	61	13.1	314	10	US-09-771-730-32
29	61	13.1	314	10	US-09-771-730-34
30	61	13.1	314	10	US-09-771-730-36
31	61	13.1	330	10	US-09-886-055-199
32	60.5	13.0	1569	9	US-10-108-605-303
33	60.5	13.0	1642	10	US-09-925-442-2
34	60.5	13.0	1648	10	US-09-925-442-35
35	60	12.8	332	9	US-09-738-626-5674
36	59	12.6	159	9	US-09-764-868-960
37	59	12.6	173	10	US-09-925-297-688
38	59	12.6	394	9	US-09-736-457-805
39	59	12.6	394	9	US-09-902-941-805
40	59	12.6	394	9	US-09-902-941-827
41	59	12.6	394	9	US-09-849-626-805
42	59	12.6	394	9	US-09-849-626-827
43	59	12.6	401	9	US-09-902-941-1917
44	59	12.6	401	9	US-09-849-626-1917
45	59	12.6	401	9	US-09-849-626-1917

ALIGNMENTS

RESULT 1
US-09-767-215-2
; Sequence 2, Application US/09767215
; Patent No. US20020081636A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; FILE REFERENCE: 07334-142001
; CURRENT APPLICATION NUMBER: US/09/767,215
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/181,159
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-767-215-2

Query Match	100.0%	Score 467;	DB 10;	Length 1004;
Best Local Similarity	100.0%;	Pred. No. 5.9e-48;		
Matches	93;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	QVTMLAFQGDALLEQISVIGNLGTGIFHRVTPGSAADQMALRPGTQIVWVDYEASEPLF	60	
Db	568	QVTMLAFQGDALLEQISVIGNLGTGIFHRVTPGSAADQMALRPGTQIVWVDYEASEPLF	627	
Qy	61	KAVLEDTTLEEAVALRRVDFCCCLSVKVTDTG	93	
Db	628	KAVLEDTTLEEAVALRRVDFCCCLSVKVTDTG	660	
RESULT 2				
US-09-767-215-5				
; Sequence 5, Application US/09767215				
; Patent No. US20020081636A1				
; GENERAL INFORMATION:				
; APPLICANT: Bertin, John				
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED				
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED				

```
; FILE REFERENCE: 07334-142001
; CURRENT APPLICATION NUMBER: US/09/767,215
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/181,159
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1138
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-767-215-5

Query Match      100.0%; Score 467; DB 10; Length 1138;
Best Local Similarity 100.0%; Pred. No. 7e-48;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QVTMLAFQGDALLEQISVIGGNLTGIFHRVTPGSAADQMALRPGTQIVMVDYEAASEPLF 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 QVTMLAFQGDALLEQISVIGGNLTGIFHRVTPGSAADQMALRPGTQIVMVDYEAASEPLF 676

Qy 61 KAVLEDTTLEAVGLLRVDGFCCLSVKVTG 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 677 KAVLEDTTLEAVGLLRVDGFCCLSVKVTG 709

RESULT 3
US-10-032-159A-14
; Sequence 14, Application US/10032159A
; Patent No. US20020164703A1
; GENERAL INFORMATION:
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,
; FILE OF INVENTION: ENCODING NUCLEIC ACIDS, AND METHODS OF USE
; FILE REFERENCE: P-LJ 5100
; CURRENT APPLICATION NUMBER: US/10/032,159A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,457
; PRIOR FILING DATE: 2000-12-21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-159A-14

Query Match      28.5%; Score 133; DB 9; Length 113;
Best Local Similarity 36.6%; Pred. No. 4.8e-09;
Matches 30; Conservative 16; Mismatches 36; Indels 0; Gaps 0;

Qy 9 GDALLEQISVIGGNLTGIFHRVTPGSAADQMALRPGTQIVMVDYEAASEPLFKAVLEDTT 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32 GDSLTSQLTLGGNARGSFVHSVKPGSLAEKAGLRGHLLEGGCIRGERQSVPLDTCT 91

Qy 69 LEEAVGLLRVDGFCCLSVKVN 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92 KEAHWTIQRCSGPVTLHYKVN 113

RESULT 4
US-10-032-159A-8
; Sequence 8, Application US/10032159A
; Patent No. US20020164703A1
; GENERAL INFORMATION:
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,
; FILE OF INVENTION: ENCODING NUCLEIC ACIDS, AND METHODS OF USE
; FILE REFERENCE: P-LJ 5100
```

```
; CURRENT APPLICATION NUMBER: US/10/032,159A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,457
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-159A-8

Query Match      28.5%; Score 133; DB 9; Length 1247;
Best Local Similarity 36.6%; Pred. No. 1.1e-07;
Matches 30; Conservative 16; Mismatches 36; Indels 0; Gaps 0;

Qy 9 GDALLEQISVIGGNLTGIFHRVTPGSAADQMALRPGTQIVMVDYEAASEPLFKAVLEDTT 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 GDSLTSQLTLGGNARGSFVHSVKPGSLAEKAGLRGHLLEGGCIRGERQSVPLDTCT 816

Qy 69 LEEAVGLLRVDGFCCLSVKVN 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 817 KEAHWTIQRCSGPVTLHYKVN 838

RESULT 5
US-09-963-959-11
; Sequence 11, Application US/09963959
; Patent No. US20020165145A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 33521, A NOVEL HUMAN GUANINE NUCLEOTIDE EXCHANGE FAMILY MEMBER
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 10448-095001
; CURRENT APPLICATION NUMBER: US/09/963,959
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 2001-09-25
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-09-963-959-11

Query Match      16.6%; Score 77.5; DB 9; Length 80;
Best Local Similarity 35.2%; Pred. No. 0.014;
Matches 25; Conservative 10; Mismatches 19; Indels 17; Gaps 3;

Qy 17 SVIGGNLTGIFHRVTPGSAADQMA-----LRPGTQIVMVDYEAASEPLFKAVLEDTTLE 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 SIVG---GIFVSSVVPSPAKAGKSGLLKVGVDVILEVNGETS-----VEGLTHE 54

Qy 71 EAVGLLRVDG 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 EAVDLLKKAG 65

RESULT 6
US-09-951-402-3
; Sequence 3, Application US/09951402
; Patent No. US20020168752A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/951,402
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: US 09/306,998
```

[illegible]

```
; PRIOR FILING DATE: 1999-05-07  
; PRIOR APPLICATION NUMBER: US 60/084,740  
; PRIOR FILING DATE: 1998-05-08  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 2037  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-922-101-3
```

Query Match 15.6%; Score 73; DB 9; Length 2037;
Best Local Similarity 31.2%; Pred.No. 3.3;
Matches 24; Conservative 14; Mismatches 25; Indels 14; Gaps 3;

QY 17 SVTGG-----NLTGFIHRV-TPGSAADOMALRPGQTIVMVDYEASEPLFRAVLDTTLLE 70
|::|| |::| |::| |::| |::| |::| |::| |:
Db 1967 SIVGGYGSPHGLPIYKTVFAKGAASEDGRKLKGDIIVANGQS-----LEGVTHE 2018
|::|| |::| |::| |::| |::| |::| |::| |:

QY 71 EAVGLLRRVDGFCCLSV 87
||| :|| | |
Db 2019 EVAAILKRKTGTITLMV 2035
||| :|| | |

RESULT 7
US-09-951-401-3
; Sequence 3, Application US/09951401
; Patent No. US2002011510A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: MMSC2 - An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/951,401
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: US 09/306,998
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 60/084,740
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-951-401-3

Query Match 15.6%; Score 73; DB 10; Length 2037;
Best Local Similarity 31.2%; Pred.No. 3.3;
Matches 24; Conservative 14; Mismatches 25; Indels 14; Gaps 3;

QY 17 SVTGG-----NLTGFIHRV-TPGSAADOMALRPGQTIVMVDYEASEPLFRAVLDTTLLE 70
|::|| |::| |::| |::| |::| |::| |::| |:
Db 1967 SIVGGYGSPHGLPIYKTVFAKGAASEDGRKLKGDIIVANGQS-----LEGVTHE 2018
|::|| |::| |::| |::| |::| |::| |::| |:

QY 71 EAVGLLRRVDGFCCLSV 87
||| :|| | |
Db 2019 EVAAILKRKTGTITLMV 2035
||| :|| | |

RESULT 8
US-09-922-101-3
; Sequence 3, Application US/09922101
; Patent No. US20020146711A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: MMSC2 - An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/922,101
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 09/306,998
; PRIOR FILING DATE: 1999-05-07

```

; NAME/KEY: SITE
; LOCATION: (13)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (148)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (149)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-948

Query Match          15.0%; Score 70; DB 9; Length 158;
Best Local Similarity 41.2%; Pred. No. 0.27;
Matches 14; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

Qy      16 ISVIGGNLTGFIHRTVPGSAAQMALRPGTQIV 49
       |::| |::| ::|| | | | | | | | | | | |
Db      102 ISIVSGKGGIYVKVTGVGSIAHQAGLEYGDQLL 135

RESULT 13
US-09-911-826A-16
; Sequence 16, Application US/09911826A
; Patent No. US20020143164A1
; GENERAL INFORMATION:
; APPLICANT: Rotin, Daniela and Pham, Nam
; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: DDW-5001-US
; CURRENT APPLICATION NUMBER: US/09/911.826A
; CURRENT FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: PCT/CA00/00042
; PRIOR FILING DATE: 2000-01-20
; PRIOR APPLICATION NUMBER: 2,259,830
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-911-826A-16

Query Match          14.7%; Score 68.5; DB 10; Length 77;
Best Local Similarity 30.8%; Pred. No. 0.16;
Matches 20; Conservative 13; Mismatches 23; Indels 9; Gaps 2;

Qy      13 LEQISVIGGN-LTGIFIHRTVPGSAAQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEE 71
       |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db      20 LRGAIAATGNAAGVIISHVEPGSKAQDVGLKRQGIHEVNGQS-----LDHWTSKR 71

Qy      72 AVGLL 76
       | : |
Db      72 ALEIL 76

RESULT 14
US-09-764-868-964
; Sequence 964, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764.868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 964
```

Query Match 14.2%; Score 66.5; DB 9; Length 316;
Best Local Similarity 32.5%; Pred. NO. 1.8;
Matches 27; Conservative 12; Mismatches 35; Indels 9; Gaps 3;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 7.125 Seconds
(without alignments)
1254.807 Million cell updates/sec

Title: US-09-767-215-2_COPY_568_660
 perfect score: 467
 Sequence: 1 QVTMLAFQGDALLEOISVIG.....GLRRYVDGFCCTSVKVNTDG 93

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 2832224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database :      PIR_73:*
1:  pir1:
2:  pir2:
3:  pir3:
4:  pir4:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %			DB	ID	Description
		Match	Length				
1	85.5	18.3	1367	2	T13703	tamA protein - fru	
2	81	17.3	1163	2	JE0366	tight junction pro	
3	77	16.5	775	2	I48236	tight junction pro	
4	77	16.5	1116	2	I54378	gene X104 protein	
5	76.5	16.4	853	2	T29736	hypothetical prote	
6	76	16.3	2054	2	T46612	multi PDZ domain p	
7	74.5	16.0	538	2	I59291	betal-syntrophin -	
8	74.5	16.0	2450	2	S71625	protein-tyrosine-p	
9	73.5	15.7	2055	2	T03259	multiple PDZ domain	
10	72	15.4	387	2	G97295	periplasmic trypsi	
11	72	15.4	717	2	T33295	hypothetical prote	
12	71.5	15.3	494	2	S27696	tcnM protein - Str	
13	71.5	15.3	503	2	I50791	hypothetical prote	
14	71.5	15.3	2294	2	I67630	hypothetical prote	
15	71.5	15.3	2466	2	I67629	protein-tyrosine-p	
16	71.5	15.3	2490	1	A54971	protein-tyrosine-p	
17	71	15.2	2172	2	T20145	hypothetical prote	
18	70	15.0	1136	2	AB1581	different protein	
19	70	15.0	1281	2	T00346	hypothetical prote	
20	68.5	14.7	394	2	S74643	proteinase hmoA (E	
21	68.5	14.7	408	1	E24409	biphenyl dioxygena	
22	68.5	14.7	408	1	F41858	biphenyl dioxygena	
23	68.5	14.7	782	2	T48246	ribonuclease II-li	
24	68.5	14.7	823	2	AF3361	endopeptidase La (
25	68	14.6	352	2	T22159	hypothetical prote	
26	68	14.6	723	2	T14765	hypothetical prote	
27	67.5	14.5	505	2	S62894	alpha-syntrophin -	
28	67	14.3	295	2	E69398	3-hydroxyacyl-CoA	
29	67	14.3	362	2	AG2118	serie/threonine k	

ALIGNMENTS

RESULT 1

Tl3703

tamA protein - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: Tl3703
E:Takahisa, M.; Togashi, S.; Suzuki, T.; Kobayashi, M.; Murayama, A.; Kondo, K.; Miya
Genes Dev. 10, 1783-1795, 1996
A:Title: The *Drosophila* tamou gene, a component of the activating pathway of extramac
A:Reference number: Z17700; MUID:96312452; PMID:8698238
A:Accession: Tl3703
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1367 <TA>
A:Cross-references: EMBL:D83477; NID:gl498136; PIDN:BA11923.1; PID:gl498137
C:Genetics:
A:Gene: tamou (tam)
A:Cross-references: FlyBase:FBgn0003177

```

Query Match      18.3%; Score 85.5; DB 2; Length 1367;
Best Local Similarity 33.8%; Pred. No. 0.63;
Matches 27; Conservative 11; Mismatches 29; Indels 13; Gaps 3;

QY  5 LAFGCDALLEQISVIGGNLTGFTHRTVPGSAQDMALRPGTQTIVMVDYEASEPLFAVL 64
    : : | : | : | | | | | | | | | | | | : :
Db   409 ISFOKEGSV-GIRLTKNEAGIFVTAYQPSPASLOGLMFGDKILKN-----DMDM 459

QY          65 EDTTLEEAV----GLLRVD 80
                | | | | |
Db          460 NGVTRREAVILFLLSLODRID 479

```

RESULT 2
JE0366
tight junction protein, ZO-2 - chicken
C:Species: Gallus gallus (chicken)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 21-Jul-2000
C:Accession: JE0366
R:Collins, J.R.; Rizzolo, L.J.
Biochem. Biophys. Res. Commun. 252, 617-622, 1998
A:Title: Protein-binding domains of the tight junction protein, ZO-2, are highly cons
A:Reference number: JE0366; MUID:99057550; PMID:9837755
A:Accession: JE0366
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1163 <COL>
A:Cross-references: GB|AF085184; NID:q3820579; PID:AAC95469.1; PID:q3820580

Query Match 17.3%; Score 81; DB 2; Length 1163;
Best Local Similarity 35.3%; Pred. No. 1.6;
Matches 24; Conservative 16; Mismatches 20; Indels 8; Gaps 4;

C:Species: Homo sapiens (man)
C>Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 21-Jul-2000
C:Accession: I59291
R:Ahm, A.H.; Yoshida, M.; Anderson, M.S.; Feener, C.A.; Selig, S.; Hagiwara, Y.; Ozawa, Proc. Natl. Acad. Sci. U.S.A. 91, 4446-4450, 1994
A:Title: Cloning of human basic AL, a distinct 59-kDa dystrophin-associated protein encoded by the dystrophin gene
A:Reference number: I59291; MUID:94240154; PMID:8183929
A:Accession: I59291
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-538 <RES>
A:Cross-references: GB:L31529; NID:g1066339; PIDN:AA81523.1; PID:g1066340
C:Genetics:
A:Gene: GDB:SWT2B1; 59-DAP
A:Cross-references: GDB:371344; OMIM:600026
A:Map position: 8q23-8q24
C:Superfamily: GLGF domain homology
F:118-191/Domain: GLGF domain homology <GLG3>

Query Match 16.0%; Score 74.5; DB 2; Length 538;
Best Local Similarity 35.6%; Pred. No. 3.4;
Matches 32; Conservative 10; Mismatches 31; Indels 17; Gaps 4;

QY 8 OGDALLEQ-----ISVIGG--NLTGIFHRVTPGSAADQ-MALRPGTQIVWVDYEASEP 58
Db 111 RGVKVLQELGGIGISIKGKKNKMPILSIKIFKGLAQTQALYVGDAILSVN----- 164
QY 59 LFKAVLEDTTLEEAVALLRVVDGFCCLSVK 88
Db 165 --GADLRDATHDEAVQALKRAGREVLEVK 192

RESULT 8
S71625
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 13 - mouse
N:Alternate names: epidermal growth factor-binding protein; serine proteinase
C:Species: Mus musculus (house mouse)
C>Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 21-Jun-2002
C:Accession: S71625; PID:g1232103; NID:g1232103; PIDN:BAA12158.1; PID:g1232104
R:Chida, D.; Kume, T.; Mukouyama, Y.; Tabata, S.; Nomura, N.; Thomas, M.L.; Watanabe, T. FEBS Lett. 358, 233-239, 1995
A:Title: Characterization of a protein tyrosine phosphatase (RTP) expressed at a very early stage of embryonic development
A:Reference number: S71625; MUID:95145716; PMID:7843407
A:Accession: S71625
A:Molecule type: mRNA
A:Residues: 1-2450 <CHI>
A:Cross-references: EMBL:D83966; NID:g1232103; PIDN:BAA12158.1; PID:g1232104
A:Experimental source: strain DBA/2; cell line MEL 745A
R:Wolf, B.B.; Brown, M.D.
FEBS Lett. 376, 177-180, 1995
A:Title: Epidermal growth factor-binding protein activates soluble and receptor-bound src kinase
A:Reference number: S67987; MUID:96105375; PMID:7498536
A:Accession: S67987
A:Molecule type: protein
A:Residues: 1098-1102 <WOL>
A:Experimental source: submaxillary glands
R:Sato, T.; Irie, S.; Kitada, S.; Reed, J.C.
Science 268, 411-415, 1995
A:Title: PAP-1: a protein tyrosine phosphatase that associates with Fas.
A:Reference number: I59595; MUID:95232528; PMID:7536343
A:Accession: I81210
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1329-1354, 'K' 1356-1447, 'R' 1449-1454 <RES>
A:Cross-references: GB:L34581; NID:g806295; PIDN:AAC42055.1; PID:g806296
R:Handriks, W.; Bruggman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.
submitted to the EMBL Data Library, June 1993
A:Description: Assessment of the expression levels of murine protein-tyrosine phosphatases

[illegible]

; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: G97295
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-387 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK81154.1; PID:gl5026290; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3218
C:Superfamily: Escherichia coli trypsin-like proteinase; GLGF domain homology; trypsin h

Query Match 15.4%; Score 72; DB 2; Length 387;
Best Local Similarity 23.7%; Pred. No. 4.4;
Matches 22; Conservative 17; Mismatches 36; Indels 18; Gaps 3;

QY 9 GDALLEOISVIGG-----NLTGIFHRTVTPGSAADQALRPQTQIVMVVYEASEPLF 60
Db 288 GKVIKPFVIGVDIKVRSQDNMKGVYKVPFGSGAKAGLRPSDILBELNGQ----- 341
QY 61 KAVLEDTTLEAVGLLRVVDGFCCLSKVKNVDG 93
Db 342 -RLSTNDIGSIVSSKIGDKVPC---KVRNNG 370

RESULT 11
T33295
hypothetical protein F26D11.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33295
R:Geisel, C.; Bradshaw, H.
submitted to the EMBL Data Library, May 1998
A:Description: The sequence of C. elegans cosmid F26D11.
A:Reference number: Z21316
A:Accession: T33295
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-717 <GPI>
A:Cross-references: EMBL:AF068716; PIDN:AAC17752.1; GSPDB:GN00023; CESP:F26D11.11
A:Experimental source: strain Bristol N2; clone F26D11
C:Genetics:
A:Gene: CESP:F26D11.11
A:Map position: 5
A:Introns: 34/2; 119/2; 226/1; 302/3; 368/3; 576/2; 623/2

Query Match 15.4%; Score 72; DB 2; Length 717;
Best Local Similarity 29.1%; Pred. No. 9;
Matches 23; Conservative 18; Mismatches 30; Indels 8; Gaps 3;

QY 5 LAFQGDALLEOISVIGNLGTGIFHRTVTPGSAADQALRPQTQIVMVVD-----YEASEPL 59
Db 676 LSFAGGTSNDPAPNSGD-SGLFTVTKVTPGSAAYRCGLREGDKLIRANDVNMINASQDNA 684

QY 60 FKAVLEDTTLEAVGLLR 78
Db 685 MEAIKKRETVELVV--LRR 701

RESULT 12
S27696
tcmN protein - Streptomyces glaucescens
C:Species: Streptomyces glaucescens
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 22-Oct-1999
C:Accession: B42276; S05976; S27696
R:Summers, R.G.; Wendt-Pienkowski, E.; Motamedi, H.; Hutchinson, C.R.
J. Bacteriol. 174, 1810-1820, 1992
A:Title: Nucleotide sequence of the tcmII-tcmIV region of the tetracenomycin C biosynthetic
dehydratase-O-methyl transferase.
A:Reference number: A42276; MUID:92193265; PMID:1548230
A:Accession: B42276

A:Status: preliminary
A:Molecule type: mRNA
A:Cross-references: EMBL:X15312

Query Match 15.3%; Score 71.5; DB 2; Length 494;
Best Local Similarity 30.4%; Pred. No. 6.6;
Matches 31; Conservative 14; Mismatches 28; Indels 29; Gaps 5;

QY 5 LAFQGDALLEOISVIGG-----NLTGIF--IHRVTPGS-----AADOM 40
Db 320 LADQGMERSRIADLGGDGWFLAQILRRHPHATGLLMDLPRVAASAGPVLEAKVADR 379
QY 41 ALRPG---TQIVMVVYEASEPLFKAVLEDTTLEAVGLLRV 79
Db 380 TVLPFGGFTDPVPTGYDAY--LFKGVLHNMSDERAVTVLRV 419

RESULT 13
T50791
hypothetical protein T30N20_90 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50791
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.;
submitted to the Protein Sequence Database, July 2000
A:Reference number: Z25240
A:Accession: T50791
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-503 <BEV>
A:Cross-references: EMBL:AL365234
A:Experimental source: cultivar Columbia; BAC clone T30N20
C:Genetics:
A:Map position: 5
A:Introns: 220/3; 255/3; 312/3
A:Note: T30N20_90

Query Match 15.3%; Score 71.5; DB 2; Length 503;
Best Local Similarity 28.4%; Pred. No. 6.7;
Matches 19; Conservative 14; Mismatches 33; Indels 1; Gaps 1;

QY 17 SVIGNLTGIFHRTVTPGSAADQALRPQTQIVMVVYEASEPLFKAVLEDTTLEAVGLL 76
Db 203 ALGVATPSGVFVHQLGPGALGVLAFTSPAT-IVILGFFIYEKRSSVTPTQTKKTDGLG 261

QY 77 RRVDDGFC 83
Db 262 VAVGMC 268

RESULT 14
I67630
protein-tyrosine-phosphatase (EC 3.1.3.48) PTPN13, nonreceptor type 13, splice form 3
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 30-Jun-2002
C:Accession: I67630
R:Maekawa, K.; Imagawa, N.; Nagamatsu, M.; Harada, S.
FEBS Lett. 337, 200-206, 1994
A:Title: Molecular cloning of a novel protein-tyrosine phosphatase containing a member
A:Reference number: I53483; MUID:94116679; PMID:8287977
A:Accession: I67630
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA

Search completed: January 22, 2003, 08:54:39
Job time : 10.125 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 3.625 Seconds
(without alignments)
1064.082 Million cell updates/sec

Title: US-09-767-215-2_COPY_568_660

Perfect score: 467

Sequence: 1 QVTMLAFQGDALLEQISVIG.....GLLRVDFGCLSVKVTNDG 93

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	467	100.0	1004	1	CARE_HUMAN
2	381	81.6	999	1	CARE_MOUSE
3	139	29.8	1147	1	CARB_HUMAN
4	80	17.1	933	1	ZO3_HUMAN
5	78	16.7	898	1	ZO3_CANFA
6	77	16.5	1174	1	ZO2_CANFA
7	77	16.5	1190	1	ZO2_HUMAN
8	75	16.1	1167	1	ZO2_MOUSE
9	73	15.6	905	1	ZO3_MOUSE
10	71.5	15.3	494	1	TCMN_STRGA
11	71.5	15.3	2485	1	PTND_HUMAN
12	68.5	14.7	408	1	BPHG_BURCE
13	68.5	14.7	812	1	LON_BRUAB
14	66.5	14.2	458	1	HRA2_HUMAN
15	66	14.1	462	1	YEL1_RALSO
16	66	14.1	692	1	FLHA_SALTY
17	66	14.1	4367	1	DYHC_NEURC
18	65.5	14.0	310	1	YE59_PYRHO
19	65.5	14.0	458	1	HRA2_MOUSE
20	65.5	14.0	552	1	AI75_HUMAN
21	65	13.9	695	1	PARE_CAUCR
22	65	13.9	710	1	RNR_THEMA
23	65	13.9	982	1	NIA_NEURC
24	64	13.7	1816	1	AF6_HUMAN
25	63.5	13.6	645	1	HRC2_XANCV
26	63	13.5	1199	1	NIFJ_ANASP
27	63	13.5	1591	1	TIAM_HUMAN
28	62.5	13.4	605	1	GLMS_THEMA
29	62.5	13.4	1461	1	PRAX_HUMAN
30	62	13.3	692	1	FLHA_YEREN
31	62	13.3	852	1	DLG2_RAT
32	61.5	13.2	443	1	YAEI_HAEIN
33	61.5	13.2	496	1	CIMA_METHH
					Q9bx16 homo sapien
					Q99xf0 mus musculu
					Q9bx17 homo sapien
					Q95049 homo sapien
					O62683 canis famil
					Q95168 canis famil
					Q9udy2 homo sapien
					Q9z0u1 mus musculu
					Q9qxy1 mus musculu
					P16559 streptomyce
					Q12923 homo sapien
					P37337 burkholderi
					O52605 brucella ab
					O43464 homo sapien
					Q8xz14 raistonia s
					P40729 salmonella
					P45443 neurospora
					O59128 pyrococcus
					Q9jiv5 mus musculu
					Q9y6n9 homo sapien
					O34479 caulobacter
					Q9wz11 thermotoga
					P08619 neurospora
					P55196 homo sapien
					P80150 xanthomonas
					Q06879 anabaena sp
					Q13009 homo sapien
					Q9wxz5 t glucosami
					Q9bxm0 homo sapien
					Q56887 versinia en
					Q93622 rattus norv
					P44936 haemophilus
					O26819 methanobact

RESULT 1
CARE_HUMAN ID CARE_HUMAN STANDARD; PRT; 1004 AA.
AC Q9BXL6; Q9BVB5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 14 (CARD-containing MAGUK protein 2) (Carna 2).
DE CARD14 OR CARMA2.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21192234; PubMed=11278692;
RA Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA Srinivasula S.M., Merriam S., Distefano P.S., Alnemri E.S.;
RT "CARD11 and CARD14 are novel caspase recruitment domain (CARD)/membrane-associated guanylate kinase (MAGUK) family members that interact with Bcl10 and activate NF-kappaB.";
RL J. Biol. Chem. 276.11877-11882(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21255663; PubMed=11356195;
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschoopp J.;
RT "Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10 phosphorylation and NF-kappaB activation.";
RL FEBS Lett. 496:121-127(2001).
RN [3]
RP ERRATUM.
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschoopp J.;
RL FEBS Lett. 505:198-198(2001).
RN [4]
RP SEQUENCE OF 1-740 FROM N.A.
RC TISSUE=Cervix, and Colon;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the phosphorylation of Bcl10.
CC -!- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD interaction.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Expressed in placenta. Also detected in HeLa S3 cells, but not in the other cancer cell lines tested.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- CAUTION: Supposed to contain a SH3 domain which is not detected by PROSITE, Pfam or SMART.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

P33147 xenopus lae
O27984 archaeoglob
P76298 escherichia
Q15700 homo sapien
Q987n5 rhizobium l
P09139 rattus norv
O34098 spiroplasma
P28191 caenorhabdi
Q09773 schizosacch
P58107 homo sapien
P32425 streptomyce
Q9pw72 gallus gall

ALIGNMENTS


```

[3]
RN ERRATUM.
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RL FEBS Lett. 505:198-198(2001).
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the
CC phosphorylation of Bcl10.
CC -!- SUBUNIT: CARD11 and Bcl10 bind to each other by CARD-CARD
CC interaction.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Detected in adult peripheral blood leukocytes,
CC thymus, spleen and liver. Also found in promyelocytic leukemia HL-
CC 60 cells, chronic myelogenous leukemia K562 cells, Burkitt's
CC lymphoma Raji cells and colorectal adenocarcinoma SW480 cells. Not
CC detected in HeLa S3, Molt-4, A549 and G431 cells.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- CAUTION: Supposed to contain a SH3 domain which is not detected by
CC PROSITE, Pfam or SMART.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF322641; AAG53402.1; -
DR Genew; HGNC:16393; CARD11.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50209; CARD; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; FALSE_NEG.
DR PROSITE; PS50106; PDZ; FALSE_NEG.
KW Coiled coil.
FT DOMAIN 11 103 CARD.
FT DOMAIN 123 442 COILED COIL (POTENTIAL).
FT DOMAIN 673 748 PDZ.
FT DOMAIN 966 1133 GUANYLATE KINASE.
FT CONFLICT 808 808 P -> L (IN REF. 2).
SQ SEQUENCE 1147 AA; 132641 MW; 913A4B015D2B36CC CRC64;

Query Match 29.8%; Score 139; DB 1; Length 1147;
Best Local Similarity 36.5%; Pred. No. 3.7e-07;
Matches 31; Conservative 17; Mismatches 37; Indels 0; Gaps 0;

QY 9 GDALLEQISVIGNLTGIFHRVTFGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTT 68
Db 666 GDSLTSQLTLGGNARGSPVHSVKPGSLAEKAGLRGHQLLLLEGCI RGERQSVPLDTCT 725

QY 69 LEEAVGLLRVDGFCCLSVKVNWDG 93
Db 726 KEEAHWTQRCSPGVTLYKVNHEG 750

RESULT 4
Z03_HUMAN
ID Z03_HUMAN STANDARD: PRT; 933 AA.
AC 095049;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona
DE occludens 3 protein) (tight junction protein 3).
GN TJP3 OR ZO3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Brower A., Gordon L., Dias J., Ramirez M.,
RA Stilwagen S., Phan H., Velasco N., Do L., Regala W., Terry A.,
RA Garnes J., Danganan L., Erler A., Christensen M., Georgescu A.,
RA Avila J., Liu S., Attix C., Andreise T., Trankheim M.,
RA Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R.,
RA Thomas P., Quan G., Kronmiller B., Arellano A., Sanders C., Ow D.,
RA Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
RT serine protease gene cluster";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-1 (BY
CC SIMILARITY).
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AC005954; AAC72274.1; ALT_INIT.
DR HSSP; P31016; 1BPE.
DR Genew; HGNC:11829; TJP3.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50106; PDZ; 3.
KW Tight junction; SH3 domain; Repeat; Membrane.
FT DOMAIN 11 93 PDZ 1.
FT DOMAIN 195 272 PDZ 2.
FT DOMAIN 394 460 PDZ 3.
FT DOMAIN 489 563 SH3.
FT DOMAIN 675 775 GUANYLATE KINASE.
SQ SEQUENCE 933 AA; 102800 MW; 680298CFD0615B47 CRC64;

Query Match 17.1%; Score 80; DB 1; Length 933;
Best Local Similarity 34.4%; Pred. No. 0.72;
Matches 21; Conservative 9; Mismatches 23; Indels 8; Gaps 1;

QY 16 ISVIGNLTGIFHRVTFGSAADQMALRPGTQIVMVDYEASEPLFKAVLEDTTLEEAVGL 75
Db 408 LRLAGENDVGFVSGVQAGSPADGGIQEGDQLQVN-----DYPFQNLTRREAVQF 459

QY 76 L 76
Db 460 L 460

RESULT 5
Z03_CANFA
ID Z03_CANFA STANDARD: PRT; 898 AA.
AC 062683;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona
DE occludens 3 protein) (tight junction protein 3).
GN TJP3 OR ZO3.

```

OS	Canis familiaris (Dog).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.	
OX	NCBI_TaxID=9615;	
RP	[1]	
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	
RX	MEDLINE=98198478; PubMed=9531559;	
RA	Haskins J., Gu L., Wittchen E.S., Hibbard J., Stevenson B.R.;	
RT	"ZO-3, a novel member of the MAGUK protein family found at the tight	
RT	junction, interacts with ZO-1 and occludin.";	
RL	J. Cell Biol. 141:199-208(1998).	
CC	-!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-1.	
CC	-!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.	
CC	-!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.	
CC	-!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.	
CC	-!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
EMBL:	AF023617; AAC39177.1; ..	
DR	HSSP; P31016; 1BFF.	
DR	InterPro; IPR000619; Guanylate_kin.	
DR	InterPro; IPR001478; PDZ.	
DR	InterPro; IPR001452; SH3.	
DR	Pfam; PF00595; PDZ; 3.	
DR	Pfam; PF00625; Guanylate_kin; 1.	
DR	SMART; SM00072; GuK; 1.	
DR	SMART; SM00228; PDZ; 3.	
DR	SMART; SM00326; SH3; 1.	
DR	PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.	
DR	PROSITE; PS00502; GUANYLATE_KINASE_2; 1.	
DR	PROSITE; PS50106; SH3; 1.	
DR	PROSITE; PS50002; SH3; 1.	
KW	Tight junction; SH3 domain; Repeat; Membrane.	
FT	DOMAIN 11 93 PDZ 1.	
FT	DOMAIN 187 264 PDZ 2.	
FT	DOMAIN 369 435 PDZ 3.	
FT	DOMAIN 464 538 SH3.	
FT	DOMAIN 650 750 GUANYLATE KINASE.	
SQ	SEQUENCE 898 AA; 984114 MW; 8091D6132DB9F15D CRC64;	
Query Match	16.7%; Score 78; DB 1; Length 898;	
Best Local Similarity	34.4%; Pred. No. 1.1; Indels 8; Gaps 1;	
Matches	21; Conservative 8; Mismatches 24;	
QY	16 ISVIGGNLTGIFHRVTPGSAADQMALRPGTOIVMYDYEAASEPLFKAVLEDTTLEAVGL 75	
DB	383 LRLAGGNDGVIFVSGVQEGSPADGGIQEGDQILQVN-----DVFFRNLTREEAVQF 434	
QY	76 L 76	
DB	435 L 435	
RESULT 6		
ID	ZO2_CANFA STANDARD; PRT; 1174 AA.	
AC	Q95168;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Tight junction protein ZO-2 (zonula occludens 2 protein) (Zona	
DE	occludens 2 protein) (Tight junction protein 2).	
GN	TJP2 OR ZO2.	
OS	Canis familiaris (Dog).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.	

OX	NCBI_TaxID=9615;	
RP	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=96421547; PubMed=8824195;	
RA	Beatch M., Jesaitis L.A., Gallin W., Goodenough D.A., Stevenson B.R.;	
RT	"The tight junction protein ZO-2 contains three PDZ (PSD-95/Discs-	
RT	Large/ZO-1) domains and an alternatively spliced region.";	
RL	J. Biol. Chem. 271:25723-25726(1996).	
RP	[2]	
RP	PARTIAL SEQUENCE FROM N.A.	
RX	MEDLINE=94179414; PubMed=8132716;	
RA	Jesaitis L.A., Goodenough D.A.;	
RT	Molecular characterization and tissue distribution of ZO-2, a tight	
RT	junction protein homologous to ZO-1 and the Drosophila discs-large	
RT	tumor suppressor protein.";	
RL	J. Cell Biol. 124:949-961(1994).	
CC	-!- FUNCTION: PLAYS A ROLE IN TIGHT JUNCTIONS AND ADHERENS JUNCTIONS.	
CC	-!- SUBUNIT: INTERACTS WITH OCCLUDIN.	
CC	-!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.	
CC	-!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE	
CC	PRODUCED BY ALTERNATIVE SPLICING.	
CC	-!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.	
CC	-!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.	
CC	-!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.	
CC	-!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
EMBL:	L27152; AAC37332.1; ..	
DR	HSSP; P31016; 1BE9.	
DR	InterPro; IPR000619; Guanylate_kin.	
DR	InterPro; IPR001478; PDZ.	
DR	InterPro; IPR001452; SH3.	
DR	Pfam; PF00018; SH3; 1.	
DR	Pfam; PF00595; PDZ; 3.	
DR	Pfam; PF00625; Guanylate_kin; 1.	
DR	SMART; SM00072; GuK; 1.	
DR	SMART; SM00228; PDZ; 3.	
DR	SMART; SM00326; SH3; 1.	
DR	PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.	
DR	PROSITE; PS50052; GUANYLATE_KINASE_2; 1.	
DR	PROSITE; PS50106; PDZ; 3.	
DR	PROSITE; PS50002; SH3; 1.	
KW	Tight junction; SH3 domain; Repeat; Membrane; Alternative splicing.	
FT	DOMAIN 10 97 PDZ 1.	
FT	DOMAIN 291 369 PDZ 2.	
FT	DOMAIN 493 574 PDZ 3.	
FT	DOMAIN 588 653 SH3.	
FT	DOMAIN 680 860 GUANYLATE KINASE.	
FT	DOMAIN 1146 1149 POLY-GLU.	
SQ	SEQUENCE 1174 AA; 132085 MW; 45AB836BDBDB1226 CRC64;	
Query Match	16.5%; Score 77; DB 1; Length 1174;	
Best Local Similarity	34.3%; Pred. No. 2;	
Matches	23; Conservative 15; Mismatches 23; Indels 6; Gaps 3;	
QY	4 MLAF-QGDALLEQISVIGGNLTGIFHRVTPGSAADQMALRPGTOIVMYDYEAASEPLFKA 62	
DB	496 MVRFRKGDV--GLRLAGGNDGVIFVAGIQEGTSABEQGLQEGDQILKVNTODFRGL--- 550	
QY	63 VLEDTTTL 69	
DB	551 VREDAVL 557	
RESULT 7		
ID	ZO2_HUMAN	


```

DR EMBL: AF177526; AAD56219.2; JOINED.
DR EMBL: AF177527; AAD56219.2; JOINED.
DR EMBL: AF177528; AAD56219.2; JOINED.
DR EMBL: AF177529; AAD56219.2; JOINED.
DR EMBL: AF177532; AAD56219.2; JOINED.
DR EMBL: AF083892; AAC33121.1; -.
DR EMBL: AF083893; AAC33122.1; -.
DR EMBL: U84581; AAB41794.1; -.
DR HSSP: P31016; 1BE9.
DR Genew: HGNC:11828; TJP2.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3; 1.
DR Pfam: PF00595; PDZ; 3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR SMART: SM00072; GuKc; 1.
DR SMART: SM00228; PDZ; 3.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS50106; PDZ; 3.
DR PROSITE: PS50002; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
KW Tight junction; SH3 domain; Repeat; Membrane; Alternative splicing;
KW Alternative initiation.
FT CHAIN 1 1190
FT CHAIN 24 1190
FT INIT_MET 24 24
FT DOMAIN 33 120
FT DOMAIN 307 385
FT DOMAIN 509 590
FT DOMAIN 604 669
FT DOMAIN 690 876
FT DOMAIN 1162 1165
FT VARSPPLIC 961 1108
FT POLY-GLU.
FT CONFLICT 411 411
FT CONFLICT 782 782
FT CONFLICT 808 808
FT CONFLICT 812 814
FT CONFLICT 822 822
FT CONFLICT 829 829
FT CONFLICT 834 834
FT CONFLICT 842 842
FT CONFLICT 996 996
FT CONFLICT 1092 1095
FT CONFLICT 1136 1136
FT CONFLICT 1155 1158
FT CONFLICT 1165 1167
FT CONFLICT 1165 1167
SQ SEQUENCE 1190 AA; 133971 MW; BE2BE6F181467058 CRC64;

Query Match 16.5%; Score 77; DB 1; Length 1190;
Best Local Similarity 34.3%; Pred. No. 2;
Matches 23; Conservative 15; Mismatches 23; Indels 6; Gaps 3;

QY 4 MLAF-QGDALLEQISVIGNLGTGFIHRVTPGSAADQMALRPGTQIVMVDYEAASEPLFKA 62
DB 512 MVRFKGDSV--GLRLAGGNDVGIFVAGIQEGTSABQEGLOEQDQILKYNVTDQFRGL--- 566

QY 63 VLEDTTL 69
DB 567 VREDAVL 573

RESULT 8
Z02_MOUSE STANDARD; PRT; 1167 AA.
AC Q9Z0U1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-2 (zonula occludens 2 protein) (zona

```

```

DE occludens 2 protein) (Tight junction protein 2).
GN TJP2 OR ZO2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA;
RX MEDLINE=99150392; PubMed=10026224;
RA Itoh M., Morita K., Tsukita S.;
RT "Characterization of ZO-2 as a MAGUK family member associated with
RT tight as well as adherens junctions with a binding affinity to
RT occludin and alpha catenin.";
FL J. Biol. Chem. 274:5981-5986(1999).
CC -!- FUNCTION: PLAYS A ROLE IN TIGHT JUNCTIONS AND ADHERENS JUNCTIONS.
CC -!- SUBUNIT: INTERACTS WITH OCCLUDIN.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to a license@isb-sib.ch).
CC -----
CC EMBL: AF113005; AAD19964.1; -.
CC HSSP: P31016; 1BE9.
CC MGD: MGI:1341872; Tjp2.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3; 1.
DR Pfam: PF00595; PDZ; 3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR SMART: SM00072; GuKc; 1.
DR SMART: SM00228; PDZ; 3.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS50106; PDZ; 3.
DR PROSITE: PS50002; SH3; 1.
DR Tight junction; SH3 domain; Repeat; Membrane.
KW DOMAIN 10 97
KW PDZ 1.
KW PDZ 2.
KW PDZ 3.
KW SH3.
KW GUANYLATE_KINASE.
KW POLY-GLU.
SQ SEQUENCE 1167 AA; 131614 MW; F15DA3EBC3F9434F CRC64;

Query Match 16.1%; Score 75; DB 1; Length 1167;
Best Local Similarity 34.3%; Pred. No. 3;
Matches 23; Conservative 15; Mismatches 23; Indels 6; Gaps 3;

QY 4 MLAF-QGDALLEQISVIGNLGTGFIHRVTPGSAADQMALRPGTQIVMVDYEAASEPLFKA 62
DB 492 MVRFKGDSV--GLRLPGGNDVGIFVAGIQEGTSABQEGLOEQDQILKYNVTDQFRGL--- 546

QY 63 VLEDTTL 69
DB 547 VREDAVL 553

RESULT 9
Z03_MOUSE STANDARD; PRT; 905 AA.
ID Z03_MOUSE
AC Q9QX11;

```



```
FT CONFLICT 1698 1714 KSQEDTICTMYYPQKI -> RVKKIDFVPCFTILRRK
FT CONFLICT 1797 1797 (IN REF. 4).
FT CONFLICT 1856 1857 G -> A (IN REF. 3).
FT CONFLICT 2069 2069 AA -> G (IN REF. 4).
FT CONFLICT 2069 2069 A -> S (IN REF. 4).
FT CONFLICT 2206 2210 GLLDQ -> VARS (IN REF. 4).
SQ SEQUENCE 2485 AA; 276903 MW; 8D1B31597C66962B CRC64;

Query Match 15.3%; Score 71.5; DB 1; Length 2485;
Best Local Similarity 28.6%; Pred. No. 18;
Matches 24; Conservative 15; Mismatches 30; Indels 15; Gaps 4;

Oy 1 QVTMLAFQGDALLE-QISVIGNL-----TGIFHRTVPGSAAD-QMALRGTQIVMVD- 52
::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1090 EITLVNLRKDKAKYGLGIFGKMGRLDLGTFISSVAPGPGADLGGCLKPGDRLISVNS 1149
::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 53 -----YEASEPLFKAVLEDTTL 69
::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1150 VSLGVSHAAIEILQNAPEVDTL 1173
::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
BPHG_BURCE
ID BPHG_BURCE STANDARD; PRT; 408 AA.
AC P37337;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Biphenyl dioxygenase system ferredoxin--NAD(+) reductase component
DE (EC 1.18.1.3).
GN BPHG.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LB400;
RX MEDLINE=92234948; PubMed=1569021;
RA Erickson B.D., Mondello F.J.;
RT "Nucleotide sequencing and transcriptional mapping of the genes
RT encoding biphenyl dioxygenase, a multicomponent
RT polychlorinated-biphenyl-degrading enzyme in Pseudomonas strain
RT LB400."
RL J. Bacteriol. 174:2903-2912(1992).
RN [2]
RP REVISIONS.
RA Erickson B.D., Mondello F.J.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PART OF THE ELECTRON TRANSFER COMPONENT OF BIPHENYL
CC DIOXYGENASE, TRANSFERS ELECTRONS FROM FERREDOXIN (BPHF) TO
CC NADH.
CC -!- CATALYTIC ACTIVITY: Reduced ferredoxin + NAD(+) = oxidized
CC ferredoxin + NADH.
CC -!- COFACTOR: FAD.
CC -!- PATHWAY: Biphenyl and polychlorinated biphenyl degradation.
CC -!- SUBUNIT: THIS DIOXYGENASE SYSTEM CONSISTS OF FOUR PROTEINS: THE
CC TWO SUBUNITS OF THE HYDROXYLASE COMPONENT (BPHA AND BPHE), A
CC FERREDOXIN (BPHF) AND A FERREDOXIN REDUCTASE (BPHG).
CC -!- SIMILARITY: TO OTHER BACTERIAL RING-HYDROXYLATING DIOXYGENASE
CC FERREDOXIN REDUCTASE COMPONENTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M86348; AAB63429.1; -.
DR PIR; F41858; F41858.
DR InterPro; IPR001327; FAD_pyr_redox.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF00070; pyr_redox; 1.
DR PRINTS; PR00368; FADPNR.
DR ProDom; PD000139; FAD_pyr_redox; 1.
KW Aromatic hydrocarbons catabolism; Flavoprotein; FAD; NAD;
KW Oxidoreductase.
FT NP_BIND 4 35 FAD (ADP PART) (POTENTIAL).
FT NP_BIND 145 173 NAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 408 AA; 42953 MW; 8A52BB01688667A9 CRC64;

Query Match 14.7%; Score 68.5; DB 1; Length 408;
Best Local Similarity 31.6%; Pred. No. 5.1;
Matches 24; Conservative 12; Mismatches 35; Indels 5; Gaps 1;

Oy 11 ALLEQISVIGNLGTGIFHRTVPGSAADOMALRPTQIV----MVDYEASEPLFKAVLE 65
::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 109 ARARMAIRGGDLAGIHTLRDLADSQLRQALQPGQSLIVGGGLIGCEVATTARKLSVH 168
::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 66 DTTLEEAVGLLRVDG 81
::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 VTILEAGDELLVRVLG 184
::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
LON_BRUAB
ID LON_BRUAB STANDARD; PRT; 812 AA.
AC O52605;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-dependent protease La (EC 3.4.21.53).
GN LON.
OS Brucella abortus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=235;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2308;
RA Robertson G.T., Kovach M.E., Allen C., Ficht T.A., Roop R.M. II;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DEGRADES SHORT-LIVED REGULATORY AND ABNORMAL PROTEINS IN
CC PRESENCE OF ATP. HYDROLYZES TWO ATPs FOR EACH PEPTIDE BOND CLEAVED
CC IN THE PROTEIN SUBSTRATE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,
CC casein and denaturated serum albumin, in presence of ATP.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF042348; AAB97420.1; -.
DR MEROPS; S16.001; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_centre.
DR InterPro; IPR003111; LON.
DR InterPro; IPR001984; Lon_endopep.
DR InterPro; IPR004815; Lon_fam.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF02190; LON; 1.
DR PRINTS; PR00830; ENDOLAPTASE.
DR SMART; SM00382; AAA; 1.
DR SMART; SM00484; LON; 1.
DR TIGRFAMs; TIGR00763; lon; 1.
DR PROSITE; PS01046; LON_SER; 1.
KW Hydrolase; Serine protease; ATP-binding.
```

```
FT NP_BIND 367 374 ATP (POTENTIAL).
FT ACT_SITE 689 689 BY SIMILARITY.
SQ SEQUENCE 812 AA; 89859 MW; B9D07E9F0D7DFDBB CRC64;

Query Match 14.7%; Score 68.5; DB 1; Length 812;
Best Local Similarity 22.6%; Pred. No. 11;
Matches 24; Conservative 20; Mismatches 33; Indels 29; Gaps 3;

QY 13 LEQTSVIGNLTG-----IFTHRVTPGSAADOMALRPTQTVVWDYEA-- 55
Db 4 IEQKTPVGGSETGGADGLYAVLPLRDIVFPHMIVPLFVGREKSTRALAEVGVGDKQILL 63
QY 56 -----SEPLFAVLDETTLEAAGLLRRVDGFCCLSVKVNITDG 93
Db 64 ATQKNAADDPAPDAIYEIGTIANVLQLKLPDG----TVKVLVEG 105

RESULT 14
HRA2_HUMAN
ID HRA2_HUMAN STANDARD; PRT; 458 AA.
AC O43464; O9HB24; O9POY3; Q9POY4;
DT 16-OCT-2001 (Rel. 40, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DE Serine protease HTRA2, mitochondrial precursor (EC 3.4.21.-) (High
DE temperature requirement protein A2) (Htra2) (Omi stress-regulated
DE endoprotease) (Serine proteinase OMI).
DE PRS25 OR HTRA2 OR OMI.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND MUTAGENESIS OF SER-306.
RX MEDLINE=20112822; PubMed=10644717;
RA Faccio L., Fusco C., Chen A., Martinotti S., Bonventre J.V.,
RA Zervos A.S.;
RT "Characterization of a novel human serine protease that has extensive
RT homology to bacterial heat shock endoprotease HtraA and is regulated by
RT kidney ischemia.";
RL J. Biol. Chem. 275:2581-2588(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 3 AND 4), AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=20428457; PubMed=10971580;
RA Gray C.W., Ward R.V., Karran E.H., Turconi S., Rowles A.,
RA Viglienghi D., Southan C., Barton A., Fantom K.G., West A.,
RA Savopoulos J.W., Hassan N.J., Clinkenbeard H., Hanning C.,
RA Amegadzie B., Davis J.B., Dingwall C., Livi G.P., Creasy C.L.;
RT "Characterization of human Htra2, a novel serine protease involved in
RT the mammalian cellular stress response.";
RL Eur. J. Biochem. 267:5699-5710(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Kidney;
RX MEDLINE=20453195; PubMed=10995577;
RA Faccio L., Fusco C., Viel A., Zervos A.S.;
RT "Tissue-specific splicing of Omi stress-regulated endoprotease leads
RT to an inactive protease with a modified PDZ motif.";
RL Genomics 68:343-347(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Chen F., Do A., Do T., Meisler M., Roe B.A.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE OF 134-458, INTERACTION WITH BIRC4, AND MUTAGENESIS OF
RP ALA-134.
RX MEDLINE=21468395; PubMed=11583623;
```

Suzuki Y., Imai Y., Nakayama H., Takahashi K., Takio K., Takahashi R.;
"A serine protease, Htra2, is released from the mitochondria and
interacts with XIAP, inducing cell death.";
Mol. Cell 8:613-621(2001).
[7]
CHARACTERIZATION.
MEDLINE=20334437; PubMed=10873535;
Savopoulos J.W., Carter P.S., Turconi S., Pettman G.R., Karran E.H.,
Gray C.W., Ward R.V., Jenkins O., Creasy C.L.;
"Expression, purification, and functional analysis of the human serine
protease Htra2.";
Protein Expr. Purif. 19:227-234(2000).
-!- FUNCTION: Serine protease that shows proteolytic activity against
a nonspecific substrate beta-casein. Promotes or induces cell
death either by direct binding to and inhibition of BIRC proteins
(also called inhibitor of apoptosis proteins, IAPs), leading to an
increase in caspase activity, or by a BIRC inhibition-independent,
caspase-independent and serine protease activity-dependent
mechanism. Isoform 2 seems to be proteolytically inactive.
-!- SUBUNIT: Interacts with Mxi2. The mature protein, but not the
precursor, binds to BIRC2, BIRC3 and BIRC4/XIAP.
-!- SUBCELLULAR LOCATION: Mitochondrial, predominantly present in the
intermembrane space. Released into the cytosol following apoptotic
stimuli, such as UV treatment, and stimulation of mitochondria
with caspase-8 truncated BID/tBID.
-!- ALTERNATIVE PRODUCTS: 4 isoforms; 1/13b (shown here), 2/D-Omi,
3/p7 and 4/p4; are produced by alternative splicing.
-!- TISSUE SPECIFICITY: Isoform 1 is ubiquitous; isoform 2 is
expressed predominantly in the kidney, colon and thyroid.
-!- DOMAIN: The PDZ domain mediates interaction with Mxi2.
-!- PTM: Autoproteolytically activated.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
-!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF020760; AAB94569.2; -
EMBL; AF141305; AAF66596.1; -
EMBL; AF141306; AAF66597.1; -
EMBL; AF141307; AAF66598.1; -
EMBL; AF184911; AAG13126.1; -
EMBL; AC006544; -; NOT_ANNOTATED_CDS.
EMBL; BC000096; AAH00096.1; -
MEROPS; S01.278; -
GeneW; HGNC:14348; PRSS25.
MIM; 606441; -
InterPro; IPR001478; PDZ.
InterPro; IPR001940; Protease2C.
InterPro; IPR001254; Ser_protease_Try.
Pfam; PF00089; trypsin; 1.
Pfam; PF00593; PDZ; 1.
PRINTS; PR00834; PROTEASES2C.
SMART; SM00228; PDZ; 1.
PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Zymogen; Mitochondrion;
KW Transit peptide; Transmembrane; Apoptosis; Alternative splicing.
FT TRANSIT 1 31 MITOCHONDRION.
FT PROPEP 32 133
FT CHAIN 134 458 SERINE PROTEASE HTRA2.
FT TRANSMEM 105 125 POTENTIAL.
FT DOMAIN 166 342 SERINE PROTEASE.
FT DOMAIN 364 445 PDZ.
FT SITE 134 137 IAP-BINDING MOTIF.
FT ACT_SITE 198 198 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 228 228 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 306 306 CHARGE RELAY SYSTEM.
FT VARSPIC 238 302 MISSING (IN ISOFORM 2).

```

FT VARSPLIC 313 313 L -> LARELGAVSLQ (IN ISOFORM 3).
FT VARSPLIC 372 403 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT VARSPLIC 314 458 DGEIVGNTMKVTAGISFAIPSDRLREFLHRGKKNSSGI
FT SRSORRYGVNMLTSPSILAELQRLRPSFPDVOHGVLIHK
FT VILGSPAHRAGLRPGDVLIAIGEOMVQNAEDVYEAVRTQSQ
FT LAVQIRGREFTLTIYVTEPE -> VSETSFUPRIAPGQ
FT CGKGRFELIOGCLVFLSSLLAISQYPTSRPQHLLVLLFG
FT CPHPLLEV (IN ISOFORM 4).
FT MUTAGEN 134 134 A->M: LOSS OF INTERACTION WITH BIRC4;
FT LOSS OF INHIBITION OF BIRC4 ACTIVITY.
FT MUTAGEN 134 134 A->M: LOSS OF INTERACTION WITH BIRC4 AND
FT OF INHIBITION OF BIRC4 ACTIVITY.
FT MUTAGEN 306 306 S->A: LOSS OF PROTEASE ACTIVITY.
SQ SEQUENCE 458 AA; 48840 MW; CEA955A7D0DD8COD CRC64;

Query Match 14.2%; Score 66.5; DB 1; Length 458;
Best Local Similarity 33.3%; Pred. No. 9.6;
Matches 16; Conservative 11; Mismatches 18; Indels 3; Gaps 1;

QY 25 GIFHRTVPGSAADQMALRPGTQIVMVD---YEASEPLFKAVLEDTTL 69
I: ||| ||| : |||| I: : : |||| :
DB 350 GVLIHKVILGSPAHRAGLRPGDVLIAIGEOMVQNAEDVYEAVRTQSQL 437

RESULT 15
YE11_RALSO
ID YE11_RALSO STANDARD; PRT; 462 AA.
AC Q8XZ14;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc metalloprotease Rsc1411 (EC 3.4.24.-).
GN RSC1411 OR RS05281.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangelot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RA "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -!- COPACTOR: Zinc (Probable).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50B.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL646064; CADI5113.1; -.
CC MEROPS: M50.004; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR004387; Zn_Mprotease.
DR InterPro: IPR000130; Zn_Mtpeptdse.
DR Pfam: PF00595; PDZ; 1.
DR SMART: SM00228; PDZ; 1.
DR TIGRFAMs: TIGR00054; mem_zinc_metalprot; 1.
DR PROSITE: PS0106; PDZ; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.

```

```

KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
KW Inner membrane; Complete proteome.
FT METAL 18 18 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 19 19 POTENTIAL.
FT METAL 22 22 ZINC (CATALYTIC) (POTENTIAL).
FT TRANSMEM 102 124 POTENTIAL.
FT DOMAIN 201 283 PDZ.
SQ SEQUENCE 462 AA; 49613 MW; 3F2818F8B0C1121B CRC64;

Query Match 14.1%; Score 66; DB 1; Length 462;
Best Local Similarity 37.7%; Pred. No. 11;
Matches 20; Conservative 10; Mismatches 15; Indels 8; Gaps 3;

QY 12 LLEQIS--VIGGNLTGIFHRTVPGSAADQMALRPGTQIVMVDYEASEPLFKA 62
::||: ::||: | | |||| :: || |||| | :| :|
DB 218 VIEQVGLRLGSPVT---IAEVLPGSAGERAGLRGDDQIVRE---AGQPADQA 264

Search completed: January 22, 2003, 08:53:35
Job time : 6.625 secs

```

THIS PAGE BLANK (USPTO)


```
ID Q9BK2 PRELIMINARY; PRT; 1695 AA.
AC Q9BK2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Tight junction protein ZO-1.
GN ZO-1.
OS Hydra attenuata (Hydra) (Hydra vulgaris).
OC Eukaryota; Metazoa; Chlidaria; Hydrozoa; Hydroida; Anthomedusae;
OC Hydridae; Hydra.
OX NCBI_TaxID=6087;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21025768; PubMed=11225567;
RA Fei K., Yan L., Zhang J., Sarais M.P. Jr.;
RT "Molecular and biological characterization of a zonula occludens-1
RT homologue in Hydra vulgaris, named HZO-1."
RL Dev. Genes Evol. 210:611-616(2000).
DR EMBL; AF230482; AAK28322.1; -.
DR HSSP; P29476; IOAV.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR000906; ZU5.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00791; ZU5; 1.
DR SMART; SM00072; Gukc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00218; ZU5; 1.
DR PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
DR PROSITE; PS0002; SH3; 1.
DR PROSITE; PS0002; SH3; 1.
SQ SEQUENCE 1695 AA; 190878 MW; AE0E7D09007316F1 CRC64;

Query Match 22.3%; Score 104; DB 5; Length 1695;
Best Local Similarity 40.6%; Pred. No. 0.024;
Matches 26; Conservative 9; Mismatches 15; Indels 14; Gaps 2;

QY 16 ISVIGNLTGTFIHRVTPGSAADQALRPGTQIVM---VDYEASEPLFKAVLEDTLEE 72
Db 686 IQVAGNSVGFVAIRPSAAKEGLRPGDQIIMCNEIDF-----ENITREEA 734
QY 73 VGLL 76
Db 735 VLIL 738

RESULT 3
Q960N4 PRELIMINARY; PRT; 974 AA.
ID Q960N4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE LD43161p.
GN PYD OR CG9763.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY051965; AAK93389.1; -.
```

```
DR FlyBase; FBgn0003177; pyd.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 2.
DR PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
SQ SEQUENCE 974 AA; 107338 MW; 5FA3A7D7D0922C2E CRC64;

Query Match 18.3%; Score 85.5; DB 5; Length 974;
Best Local Similarity 33.8%; Pred. No. 1.3;
Matches 27; Conservative 11; Mismatches 29; Indels 13; Gaps 3;

QY 5 LAFQGDALLEISVIGNLTGTFIHRVTPGSAADQALRPGTQIVMVDYEASEPLFKAVL 64
Db 580 ISFQKEGVS-GIRLTGCGNEAGIFVTAVQPGSPASQGLMPGDKILKVN-----DMDM 630
QY 65 EDTTLEEA-----GLLRVD 80
Db 631 NGVTREAEVLFLLSLQDRID 650

RESULT 4
Q94880 PRELIMINARY; PRT; 1367 AA.
ID Q94880;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE TAMA.
GN PYD OR TAMOU OR CG9763.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CANTON-S;
RX MEDLINE=96312452; PubMed=8698238;
RA Takahisa M., Togashi S., Suzuki T., Kobayashi M., Murayama A.,
RA Kondo K., Miyake T., Ueda R.;
RT "The Drosophila tamou gene, a component of the activating pathway of
RT extramacrochaetae expression, encodes a protein homologous to
RT mammalian cell-cell junction-associated protein ZO-1."
RL Genes Dev. 10:1783-1795(1996).
DR EMBL; D83477; BAA11923.1; -.
DR HSSP; Q12923; 3PDZ.
DR FlyBase; FBgn0003177; pyd.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00072; Gukc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
SQ SEQUENCE 1367 AA; 148347 MW; 571C4566C6868BF8 CRC64;

Query Match 18.3%; Score 85.5; DB 5; Length 1367;
Best Local Similarity 33.8%; Pred. No. 1.9;
Matches 27; Conservative 11; Mismatches 29; Indels 13; Gaps 3;

QY 5 LAFQGDALLEISVIGNLTGTFIHRVTPGSAADQALRPGTQIVMVDYEASEPLFKAVL 64
Db 409 ISFQKEGVS-GIRLTGCGNEAGIFVTAVQPGSPASQGLMPGDKILKVN-----DMDM 459
QY 65 EDTTLEEA-----GLLRVD 80
Db 460 NGVTREAEVLFLLSLQDRID 479
```

DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PD2; 6.


```
DR HSP: P31016; 1BE9.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR Pfam: PF00595; PDZ; 3.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR01597; ZONOCCLUDNS.
DR PRINTS: PR01599; ZONOCCLUDNS2.
DR PRINTS: PR01600; ZONOCCLUDNS3.
DR SMART: SM00072; GUKC; 1.
DR SMART: SM00228; PDZ; 3.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00052; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS50106; PDZ; 3.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 1163 AA; 130668 MW; 3639431A560B1D4E CRC64;

Query Match 17.3%; Score 81; DB 13; Length 1163;
Best Local Similarity 35.3%; Pred. No. 4.9;
Matches 24; Conservative 16; Mismatches 20; Indels 8; Gaps 4;

QY 4 MLAF-QGDALLEQISVIGNLTGIFHRTVPGSAADQMALRPGTQIVMVDYEAASEPLFKA 62
DB 481 MVRFKGDSV--GLRLAGNDVGIFGTQIGETSADQGLGQDQILKVNADQ---FRG 534
QY 63 VL-EDTTL 69
DB 535 IVREDAVL 542

RESULT 10
Q96KB4 PRELIMINARY; PRT; 548 AA.
AC Q96KB4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CDNA FLJ14386 fis, clone HEMBA1002417, moderately similar to TIGHT
DE junction protein ZO-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiyo S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027292; BAB55020.1; -.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR Pfam: PF00595; PDZ; 1.
DR PRINTS: PR01597; ZONOCCLUDNS.
DR PRINTS: PR01600; ZONOCCLUDNS3.
DR PROSITE: PS00052; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS50106; PDZ; 1.
DR PROSITE: PS50002; SH3; 1.
SQ SEQUENCE 548 AA; 61005 MW; ALA18ECFDE2627AC CRC64;

Query Match 17.1%; Score 80; DB 4; Length 548;
Best Local Similarity 34.4%; Pred. No. 2.6;
Matches 21; Conservative 9; Mismatches 23; Indels 8; Gaps 1;

QY 16 ISVIGNLTGIFHRTVPGSAADQMALRPGTQIVMVDYEAASEPLFKAILEDTTLEAVGL 75

Db 23 LRLAGNDVGIFVSGVQAGSPADGGQIQEGDQILQVN-----DVPRQLTREAQVF 74
QY 76 L 76
Db 75 L 75

RESULT 11
O43386 PRELIMINARY; PRT; 579 AA.
AC O43386;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE KIA0011B (Fragment).
GN KIA0011B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Connolly K.S., Gunning K.M., Davis C.A., Kadner K., Subramanian S.,
RA Miguel T., Lewis K.D., Fridlyand J., Alcivare D., Benke J.A.,
RA Bondoc M., Bowen E., Chiang A., Critz P., Jaklevic M.A., Lindo K.,
RA Lindquist K., Miller C., Patel S., Piscia C., Riley B.E., Rojeski H.,
RA Sarmiento R., Yu C., Montenegro M., Aerts A., Chung A., Abrajano A.,
RA Baker M., Gau C., Jett J., Ko C., Beall K., Woolley J.P., Stultz J.L.,
RA Kimmerly W., Martin C.H.;
RT "Sequencing of human chromosome 5.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ricke D.O., Wagner R.P.;
RT "Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004227; AAC04379.1; -.
DR HSP: O12923; 3PDZ.
DR InterPro: IPR000595; CNMP_binding.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR000651; RasGEFN.
DR Pfam: PF00027; CNMP_binding; 1.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00618; RasGEFN; 1.
DR SMART: SM00100; CNMP; 1.
DR SMART: SM00228; PDZ; 1.
DR SMART: SM00229; RasGEFN; 1.
DR PROSITE: PS50042; CNMP_BINDING_3; 1.
DR PROSITE: PS50106; PDZ; 1.
FT NON_TER 1
FT NON_TER 579 579
SQ SEQUENCE 579 AA; 65197 MW; CAE221DE69CBF92D CRC64;

Query Match 17.0%; Score 79.5; DB 4; Length 579;
Best Local Similarity 34.6%; Pred. No. 3.1;
Matches 27; Conservative 12; Mismatches 28; Indels 11; Gaps 3;

QY 15 QISVIGNLT--GIFHRTVPGSAADQMALRPGTQIVMVDYEAASEPLFKAILEDTTLEEA 72
DB 378 QFSLNGSGSEKGFIFVEGVEPGSKAADSLKRGDQIMEVNGQN-----FENITFMKA 429
QY 73 VGLLRVDGFCCLSVKVN 90
DB 430 VEILRN-NTHLALTVKTN 446

RESULT 12
O8TEA3 PRELIMINARY; PRT; 1113 AA.
AC O8TEA3;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
```


Search completed: January 22, 2003, 08:56:42
Job time : 18.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: January 22, 2003, 08:49:35 ; Search time 13.6425 Seconds
(without alignments)
683.714 Million cell updates/sec

Title: US-09-767-215-2_COPY_676_745
Perfect score: 386
Sequence: 1 GDSFYRVNLAMEGRAKGL.....MKDTAAHGTPNYSRAQQQL 70
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	ID	Description	
1	386	100.0	1004	22	AAE07164	Human caspase recr
2	386	100.0	1139	22	AAE07165	Human predicted ca
3	146	37.8	1147	22	AAU01207	Human caspase recr
4	116	30.1	82	23	AAU73249	Human plakoglobin
5	116	30.1	350	22	AAU38934	Human polypeptide
6	116	30.1	746	22	AAU38934	Human polypeptide
7	116	30.1	1032	22	AAU01206	Human caspase recr
8	116	30.1	1032	23	AAU73247	Human plakoglobin
9	102.5	26.6	1745	23	ABU57253	Mouse ischaemic co
10	98.5	25.5	1736	23	AAU84308	Human endometrial

11	97.5	25.3	1445	22	ABB64619	Drosophila melanog
12	96.5	25.0	89	23	AAU73251	Mouse tight juncti
13	96.5	25.0	548	22	AAU93968	Human stomach canc
14	96.5	25.0	548	22	AAU93968	Human protein sequ
15	96.5	25.0	928	23	AAE21718	Human PKIN-13 prot
16	96.5	25.0	1037	22	ABG22366	Novel human diagn
17	95	24.6	85	23	AAU73253	Canine tight junct
18	95	24.6	1916	22	ABB62423	Drosophila melanog
19	95	24.6	1916	22	ABB66089	Drosophila melanog
20	94	24.4	85	23	AAU73252	Human tight juncti
21	92.5	24.0	89	23	AAU73250	Human tight juncti
22	85	22.0	92	23	AAU73254	Canine tight junct
23	78.5	20.3	674	19	AAU72748	Human p-dlg protei
24	70	18.1	77	22	ABG21078	Novel human diagn
25	69.5	18.0	400	22	AAU53113	Propionibacterium
26	65.5	17.0	194	22	ABG24921	Novel human diagn
27	63.5	16.5	724	23	ABB04805	LDL receptor bindi
28	63	16.3	156	22	ABG67213	Amino acid sequenc
29	63	16.2	465	20	AAU32501	Human parkin gene
30	62.5	16.2	104	22	ABU10653	Human pancreatic c
31	62.5	16.2	104	22	AAU92411	Human digestive sy
32	62.5	16.2	724	23	ABB04798	LDL receptor bindi
33	60	15.5	746	22	ABU67068	Drosophila melanog
34	60	15.5	765	22	ABB84456	Drosophila melanog
35	59	15.3	107	23	ABB06234	Human DNA topoisom
36	59	15.3	405	20	AAU27275	Mouse Doc2alpha pr
37	59	15.3	817	22	ABU71340	Drosophila melanog
38	58.5	15.2	80	23	ABU06301	Human ORFX protein
39	58.5	15.2	724	23	ABB04804	LDL receptor bindi
40	58	15.0	1227	19	AAU44321	Bacillus thuringie
41	58	15.0	1227	22	AAU19950	Bacillus thuringie
42	58	15.0	1228	22	AAU02039	B. thuringiensis t
43	58	15.0	1229	15	AAU54074	CryET5. Bacillus
44	58	15.0	1229	18	AAU35259	Bacillus thuringie
45	58	15.0	1229	18	AAU17699	CryET5. Bacillus

ALIGNMENTS

RESULT 1
AAE07164
ID AAE07164 standard; Protein; 1004 AA.
XX AAE07164;
AC AAE07164;
XX AAE07164;
DT 06-NOV-2001 (first entry)
XX Human caspase recruitment domain-14 (CARD-14).
DE Human; caspase recruitment domain-14; CARD-14; chromosome 17;
KW nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy;
KW autoimmune disorder; systemic lupus erythematosus; neurological disorder;
KW Alzheimer's disease; Parkinson's disease; inflammatory disorder; aneemia;
KW haematological disorder; myelodysplastic syndrome; myocardial infection;
KW stroke; immune disorder; Crohn's disease; allergic rhinitis; infection;
KW cell signalling disorder; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiviral; antibacterial.
XX Homo sapiens.

Key	Modified-site	Location/Qualifiers
FT	6..9	/note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT	Domain	10..116
FT	Modified-site	/label= CARD_domain
FT	Modified-site	12..15
FT	Modified-site	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	18..21
FT	Modified-site	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	25..27
FT	Modified-site	/note= "Protein kinase C phosphorylation site"

PT regulating growth and cell death and useful for the treatment of cancer
 _

PS Claim 1; Fig 1A-1E; 109pp; English.

XX
 CC The present sequence is human caspase recruitment domain-14 (CARD-14).
 CC The CARD-14 gene is located on chromosome 17. The CARD-14 is used for
 CC the detection of modulators that modulates the ability of CARD-14 to
 CC bind to Bcl-10 and stimulates phosphorylation of Bcl-10 or activation
 CC of nuclear factor-kappa B (NF-kB). The CARD-14 is useful for regulating
 CC growth and cell death and useful for the treatment of cancer. It is
 CC also useful for the treatment of autoimmune disorders (e.g., systemic
 CC lupus erythematosus), neurological disorders e.g., Alzheimer's and
 CC Parkinson's disease, inflammatory disorders, haematological disorders
 CC (e.g., anaemia, myelodysplastic syndromes), myocardial infarctions,
 CC strokes, immune disorders (e.g., Crohn's disease, allergic rhinitis),
 CC cell signalling disorders and certain viral and bacterial infections.

XX Sequence 1004 AA;

Query Match 100.0%; Score 386; DB 22; Length 1004;
 Best Local Similarity 100.0%; Pred. No. 7.9e-43;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDSFYIRVNLAMEGRAKGLQVHCNEVLHVTDTMFQGGCWHAHVRVNSYTMKDTAAHGTTI 60
 Db 676 GDSFYIRVNLAMEGRAKGLQVHCNEVLHVTDTMFQGGCWHAHVRVNSYTMKDTAAHGTTI 735

QY 61 PNYRAQQQL 70

Db 736 PNYRAQQQL 745

RESULT 2

AAE07165

ID AAE07165 standard; Protein; 1139 AA.

XX AC AAE07165;

XX DT 06-NOV-2001 (first entry)

XX DE Human predicted caspase recruitment domain-14 (CARD-14).

XX KW Human; caspase recruitment domain-14; CARD-14; chromosome 17;
 KW nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy;
 KW autoimmune disorder; systemic lupus erythematosus; neurological disorder;
 KW Alzheimer's disease; Parkinson's disease; inflammatory disorder; anaemia;
 KW haematological disorder; myelodysplastic syndrome; myocardial infarction;
 KW stroke; immune disorder; Crohn's disease; allergic rhinitis; infection;
 KW cell signalling disorder; cytostatic; immunosuppressive; neutropic;
 KW neuroprotective; antiviral; antibacterial.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Misc-difference 700 /note= "Encoded by TGG"

XX PN WO200159065-A2.

XX PD 16-AUG-2001.

XX PF 22-JAN-2001; 2001WO-US02087.

XX PR 09-FEB-2000; 2000US-0181159.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Bertin J;

XX DR WPI; 2001-497073/54.

XX DR N-PSDB; AAD13448.

XX

PT An isolated caspase recruitment domain polypeptide useful for
 PT regulating growth and cell death and useful for the treatment of cancer
 _

PS Disclosure; Fig 2A-2C; 109pp; English.

XX
 CC The present sequence is predicted human caspase recruitment domain-14
 CC (CARD-14). The CARD-14 gene is located on chromosome 17. The CARD-14 is
 CC used for the detection of modulators that modulates the ability of
 CC CARD-14 to bind to Bcl-10 and stimulates phosphorylation of Bcl-10 or
 CC activation of nuclear factor-kappa B (NF-kB). The CARD-14 is useful for
 CC regulating growth and cell death and useful for the treatment of cancer.
 CC It is also useful for the treatment of autoimmune disorders (e.g.,
 CC systemic lupus erythematosus), neurological disorders e.g., Alzheimer's
 CC and Parkinson's disease, inflammatory disorders, haematological disorders
 CC (e.g., anaemia, myelodysplastic syndromes), myocardial infarctions,
 CC strokes, immune disorders (e.g., Crohn's disease, allergic rhinitis),
 CC cell signalling disorders and certain viral and bacterial infections.

XX Sequence 1139 AA;

Query Match 100.0%; Score 386; DB 22; Length 1139;
 Best Local Similarity 100.0%; Pred. No. 9.2e-43;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDSFYIRVNLAMEGRAKGLQVHCNEVLHVTDTMFQGGCWHAHVRVNSYTMKDTAAHGTTI 60
 Db 726 GDSFYIRVNLAMEGRAKGLQVHCNEVLHVTDTMFQGGCWHAHVRVNSYTMKDTAAHGTTI 785

QY 61 PNYRAQQQL 70

Db 786 PNYRAQQQL 795

RESULT 3

AAU01207

ID AAU01207 standard; Protein; 1147 AA.

XX AC AAU01207;

XX DT 12-SEP-2001 (first entry)

XX DE Human caspase recruitment domain, CARD-11 polypeptide.

XX KW Human; caspase recruitment domain; CARD-11; Bcl-10; NF-kappaB;
 KW apoptosis; hyperproliferative disorder; autoimmune; neurological;
 KW inflammatory disorder; viral infection; stress-related response.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Domain 6..112 /note= "CARD domain"

FT Modified-site 7..9 /note= "Protein kinase C phosphorylation site"

FT Modified-site 7..10 /note= "Protein kinase II phosphorylation site"

FT Modified-site 100..102 /note= "Protein kinase C phosphorylation site"

FT Modified-site 100..103 /note= "Protein kinase II phosphorylation site"

FT Modified-site 105..107 /note= "Protein kinase C phosphorylation site"

FT Modified-site 106..109 /note= "Protein kinase C phosphorylation site"

FT Modified-site 130..431 /note= "CAMP- and cGMP-dependent protein kinase phosphorylation site"

FT Domain 130..431 /note= "Coiled coil domain"

FT Modified-site 162..165 /note= "Casein kinase II phosphorylation site"

FT Modified-site 168..171 /note= "Casein kinase II phosphorylation site"

FT Modified-site 175..183 /note= "Casein kinase II phosphorylation site"

FT

FT	Modified-site	/note= "Tyrosine kinase phosphorylation site"	FT	Modified-site	/note= "Protein kinase C phosphorylation site"
FT	182..185		FT	816..819	
FT	/note= "Casein kinase II phosphorylation site"		FT	/note= "Casein kinase II phosphorylation site"	
FT	189..195		FT	823..828	
FT	/note= "Tyrosine kinase phosphorylation site"		FT	/note= "N-myristoylation site"	
FT	241..244		FT	Modified-site	
FT	/note= "N-glycosylation site"		FT	847..850	
FT	243..245		FT	/note= "Casein kinase II phosphorylation site"	
FT	/note= "Protein kinase C phosphorylation site"		FT	853..858	
FT	282..285		FT	/note= "N-myristoylation site"	
FT	/note= "Amidation site"		FT	857..859	
FT	286..289		FT	/note= "Protein kinase C phosphorylation site"	
FT	/note= "Casein kinase II phosphorylation site"		FT	872..875	
FT	290..292		FT	/note= "Casein kinase II phosphorylation site"	
FT	/note= "Protein kinase C phosphorylation site"		FT	882..1147	
FT	378..381		FT	/note= "Guanylate kinase (GUK) domain"	
FT	/note= "Casein kinase II phosphorylation site"		FT	897..900	
FT	429..432		FT	/note= "Casein kinase II phosphorylation site"	
FT	/note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"		FT	917..922	
FT	459..461		FT	/note= "N-myristoylation site"	
FT	/note= "Protein kinase C phosphorylation site"		FT	926..929	
FT	471..474		FT	/note= "Casein kinase II phosphorylation site"	
FT	/note= "Casein kinase II phosphorylation site"		FT	935..937	
FT	472..475		FT	/note= "Protein kinase C phosphorylation site"	
FT	/note= "N-glycosylation site"		FT	1003..1006	
FT	476..479		FT	/note= "Casein kinase II phosphorylation site"	
FT	/note= "Casein kinase II phosphorylation site"		FT	1010..1018	
FT	508..510		FT	/note= "Tyrosine kinase phosphorylation site"	
FT	/note= "Protein kinase C phosphorylation site"		FT	1050..1055	
FT	510..513		FT	/note= "N-myristoylation site"	
FT	/note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"		FT	1088..1091	
FT	558..560		FT	/note= "Casein kinase II phosphorylation site"	
FT	/note= "Protein kinase C phosphorylation site"		FT	1120..1123	
FT	578..581		FT	/note= "Casein kinase II phosphorylation site"	
FT	/note= "Casein kinase II phosphorylation site"		XX	WO200140458-A2.	
FT	584..587		PN		
FT	/note= "N-glycosylation site"		XX	07-JUN-2001.	
FT	587..592		PD		
FT	/note= "N-myristoylation site"		XX	01-DEC-2000; 2000WO-US32716.	
FT	634..637		XX	03-DEC-1999; 99US-0168780.	
FT	/note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"		PR	18-FEB-2000; 2000US-0507533.	
FT	635..748		PR	25-FEB-2000; 2000US-0513904.	
FT	/note= "PDZ domain"		PR	10-OCT-2000; 2000US-0685791.	
FT	635..1147		XX	(MILL-) MILLENNIUM PHARM INC.	
FT	/note= "WAGUK domain"		PA	Bertin J;	
FT	638..641		PI		
FT	/note= "Glycosaminoglycan attachment site"		XX		
FT	678..683		DR	WPI; 2001-367809/38.	
FT	/note= "N-myristoylation site"		DR	N-PSDB; AAS05389.	
FT	687..689		XX		
FT	/note= "Protein kinase C phosphorylation site"		XX	Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10, CARD-11, useful as targets for therapy, as immunogens, and in screening and detection assays -	
FT	692..695		PT		
FT	/note= "Casein kinase II phosphorylation site"		PT	Claim 9; Fig 14A-14C; 145pp; English.	
FT	698..703		XX		
FT	/note= "N-myristoylation site"		PS		
FT	710..715		XX		
FT	/note= "N-myristoylation site"		CC		
FT	725..728		CC		
FT	/note= "Casein kinase II phosphorylation site"		CC		
FT	761..766		CC		
FT	/note= "N-myristoylation site"		CC		
FT	764..767		CC		
FT	/note= "Casein kinase II phosphorylation site"		CC		
FT	766..834		CC		
FT	/note= "SH3 domain"		CC		
FT	776..779		CC		
FT	/note= "N-glycosylation site"		CC		
FT	779..782		CC		
FT	/note= "Casein kinase II phosphorylation site"		CC		
FT	787..789		CC		

The present sequence represents novel human caspase recruitment domain, CARD-11. The polynucleotide encoding this sequence was isolated from a human T-cell cDNA library. Also described are novel human sequences for CARD-9 and CARD-10 (AAU01205, AAU01206) and rat CARD-9 (AAU01204). CARD-9, CARD-10 and CARD-11 interact with Bcl-10 which is thought to activate nuclear factor (NF)-kappaB and apoptosis. The sequences of the invention can be used for treating a disorder associated with abnormal levels of apoptosis by modulating the expression or activity of CARD-9, CARD-10, or CARD-11. They can be used for the treatment of hyperproliferative disorders (e.g., cancer) autoimmune disorders (e.g., systemic lupus erythematosus), neurological disorders (e.g., Alzheimer's disease), inflammatory disorders (e.g., Crohn's disease), and viral infection (e.g., HIV). The CARD polypeptide, polynucleotide and an antibody which selectively binds to CARD can be used in screening and detection assays (e.g., chromosomal

CC	mapping, tissue typing), predictive medicine (prognostic assays,
CC	monitoring clinical trials, and therapy (treatment and prophylaxis). The
CC	CARD polypeptide may be used to screen for drugs that bind to and/or
CC	modulate it. CARD sequences are potential targets for regulating
CC	inflammation, cancer, NF-kappaB signalling, stress-related response and
CC	apoptosis in human disease. A host cell containing a polynucleotide
CC	encoding CARD can be used to create transgenic animals.
XX	
SQ	Sequence 1147 AA;
	Query Match 37.8%; Score 146; DB 22; Length 1147;
	Best Local Similarity 46.5%; Pred. No. 1.9e-10;
	Matches 33; Conservative 12; Mismatches 24; Indels 2; Gaps 2;
QY	1 GDSFYIRVNLAMEGRAGK-ELQVHCNEVLHVTDTMFGCGCWHARRVNSYTMKDTAAHGT 59
Db	IIIIIIIIII: : : :IIIIIIII: : :II: :II
QY	766 GDSFYIRLNISSQLDACTMSLKDDVHVHVDITMYQDRHEWPCARVDPFTDHDLDW-GT 824
Db	60 IPNYSRAQQQL 70
	II:IIIIII I
Db	825 IPSYSRAQQQLL 835
RESULT 4	
AAU073249	ID AAU73249 standard; Protein: 82 AA.
XX	AAU73249;
XX	
DT	12-MAR-2002 (first entry)
DE	Human plakoglobin interacting protein PLA_2H12, SH3 domain.
XX	Human; plakoglobin; cytosstatic; osteopathic; dermatological; cardiant;
KW	plakoglobin related disease; skin carcinoma; acantholytic disease;
KW	basal cell carcinoma; squamous cell carcinoma; Naxos disease;
KW	extramammary Paget's disease; heart disease; skin blistering;
KW	subcorneal acantholysis; Grover's disease; Hailey-Hailey's disease;
KW	Darier's disease; ectodermal dysplasia; skin fragility syndrome.
XX	
OS	Homo sapiens.
XX	
PN	WO200185933-A2.
XX	
PD	15-NOV-2001.
XX	
PF	02-MAY-2001; 2001WO-EP04872.
XX	
PR	09-MAY-2000; 2000EP-0201668.
XX	
PA	(VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX	
PI	Van Roy F, Bonne S, Vanlandschoot A;
XX	
XX	WPI: 2002-062246/08.
XX	
PT	New polypeptide, useful for treating skin carcinoma or acantholytic
PT	disease such as Grover's and Darier's disease, comprises a protein
PT	interacting with human plakoglobin and involved in transduction of
PT	plakoglobin related signal to nucleus -
XX	
XX	Example 4; Figure 4; 98pp; English.
XX	
CC	The invention relates to an isolated plakoglobin interacting polypeptide
CC	(I). (I) is useful as a medicament and in the manufacture of a
CC	medicament for treating plakoglobin related diseases, such as skin
CC	carcinoma or an acantholytic disease, and to screen compounds that
CC	interfere with the interaction of the polypeptide with plakoglobin
CC	The plakoglobin related diseases include basal cell carcinoma, squamous
CC	cell carcinoma, extramammary Paget's disease, Naxos disease, heart
CC	diseases, skin blistering and acantholytic diseases such as subcorneal
CC	acantholysis, Grover's disease, Hailey-Hailey's disease or Darier's
CC	disease, and ectodermal dysplasia/skin fragility syndrome. AAU73245-

[illegible]

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX
SQ Sequence 350 AA;

Query Match 30.1%; Score 116; DB 22; Length 350;
Best Local Similarity 40.0%; Pred. NO. 4.8e-07;
Matches 28; Conservative 11; Mismatches 29; Indels 2; Gaps 2;

Qy 2 DSFYIRVNLAMEGRAGE-LQVHCNEVLHVTDTMFQCGCGWHAHRVNSYTMKDTAAHGTTI 60
Db 23 EPFYIRANLTLPERADPHALCVKAQEILRLVDSAYKRRQEWECTRVDPDLTLD-LDRGTV 81
Qy 61 PNYSRAQQQL 70
Db 82 PNYQRAQQLL 91

RESULT 6
AAM93822
ID AAM93822 standard; Protein; 746 AA.
XX
AC AAM93822;
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 3879.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
OS Homo sapiens.
XX
PN EPI130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 93JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
DR N-PSDB; AAK94778.
XX
PT 830 primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 3879; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5' and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 746 AA;

Query Match 30.1%; Score 116; DB 22; Length 746;
Best Local Similarity 40.0%; Pred. NO. 1.2e-06;
Matches 28; Conservative 11; Mismatches 29; Indels 2; Gaps 2;

Qy 2 DSFYIRVNLAMEGRAGE-LQVHCNEVLHVTDTMFQCGCGWHAHRVNSYTMKDTAAHGTTI 60
Db 419 EPFYIRANLTLPERADPHALCVKAQEILRLVDSAYKRRQEWECTRVDPDLTLD-LDRGTV 477
Qy 61 PNYSRAQQQL 70
Db 478 PNYQRAQQLL 487

RESULT 7
AAU01206
ID AAU01206 standard; Protein; 1032 AA.
XX
AC AAU01206;
DT 12-SEP-2001 (first entry)
XX
DE Human caspase recruitment domain, CARD-10 polypeptide.
XX
KW Human; caspase recruitment domain; CARD-10; Bcl-10; NF-kappaB;
KW apoptosis; hyperproliferative disorder; autoimmune; neurological;
KW inflammatory disorder; viral infection; stress-related response.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 15..20 /note= "N-myristoylation site"
FT Modified-site 18..21 /note= "Casein kinase II phosphorylation site"
FT Domain 23..123 /note= "CARD domain"
FT Modified-site 68..70 /note= "Protein kinase C phosphorylation site"
FT Modified-site 76..79 /note= "N-glycosylation site"
FT Modified-site 78..80 /note= "Protein kinase C phosphorylation site"
FT Modified-site 88..91 /note= "Amidation site"
FT Modified-site 112..115 /note= "Casein kinase II phosphorylation site"
FT Modified-site 113..118 /note= "N-myristoylation site"
FT Domain 147..457 /note= "Coiled coil domain"
FT Modified-site 201..207 /note= "Tyrosine kinase phosphorylation site"
FT Region 230..251 /note= "Leucine zipper homology region"
FT Modified-site 242..245 /note= "Casein kinase II phosphorylation site"
FT Modified-site 293..295 /note= "Protein kinase C phosphorylation site"
FT Modified-site 293..296 /note= "Casein kinase II phosphorylation site"
FT Modified-site 309..314 /note= "N-myristoylation site"
FT Modified-site 313..315 /note= "Protein kinase C phosphorylation site"
FT Modified-site 331..334 /note= "Casein kinase II phosphorylation site"
FT Domain 366..398 /note= "Tropomyosin domain"
FT Modified-site 412..415

KW Human; endometrial cancer; differential expression;
 KW DNA microarray; protein microarray.
 XX
 OS Homo sapiens.
 XX
 XX WO200209573-A2.
 XX
 XX 07-FEB-2002.
 XX
 XX 31-JUL-2001; 2001WO-US24104.
 XX
 XX 31-JUL-2000; 2000US-221735P.
 XX (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
 XX
 XX Mutter GL;
 XX
 DR WPI; 2002-179967/23.
 DR N-PSDB; ABK35528.
 XX
 XX Diagnosing endometrial cancer comprises determining expression of
 PT nucleic acid molecules or expression products that are differentially
 PT expressed in normal and malignant endometrium .
 XX
 XX Claim 33; Page 225-231; 233pp; English.
 XX
 CC The invention relates to diagnosing endometrial cancer in a subject
 CC suspected of having endometrial cancer comprising determining the
 CC expression of a set of nucleic acid molecules or expression products in
 CC an endometrial sample suspected of being cancerous, where the set of
 CC nucleic acid molecules comprises at least 2 nucleic acid molecules
 CC selected from 50 fully defined sequences as given in the specification.
 CC The nucleic acids are used as an array of at least 2 of the 50
 CC nucleic acids bound to a solid substrate. Also included is a solid-phase
 CC protein microarray comprising at least 2 antibodies or its antigen
 CC binding fragments, that specifically bind at least 2 different
 CC polypeptides from the 50 fully defined sequences as given in the
 CC specification, fixed to a solid substrate. The methods and arrays are
 CC useful for the diagnosis of endometrial cancer, selecting and monitoring
 CC treatment regimes and identification of lead compounds useful for the
 CC treatment of endometrial cancer. The present composition is one of 50
 CC proteins differentially expressed between cancerous and non-cancerous
 CC samples.
 XX
 SQ Sequence 1736 AA;
 Query Match 25.5%; Score 98.5; DB 23; Length 1736;
 Best Local Similarity 36.2%; Pred. No. 0.0008;
 Matches 25; Conservative 9; Mismatches 32; Indels 3; Gaps 2;
 QY 1 GDSFYIRVNLAMEGRAGKELQVHCNEVLHVDTMTFQG-CGCWHAHRVNSYTMKDTAAHGT 59
 DB 504 GDSFYIRHFEVEKESPYGLSFNKGVEFRAVDTLNGLGSLNLAIRICKNHKE--VERGI 561
 QY 60 IPNYSRAQQ 68
 DB 562 IPNKNRAEQ 570
 RESULT 11
 ABB64619
 ID ABB64619 standard; Protein; 1445 AA.
 XX
 AC ABB64619;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 20649.
 XX
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.

XX WO200171042-A2.
 PN
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-656860/75.
 DR
 DR N-PSDB; ABL08722.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions .
 PT
 PS Disclosure; SEQ ID NO 20649; 2lpp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABLI6176-ABLJ0511), expressed DNA
 CC sequences (ABLO1840-ABLI6175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 1445 AA;
 SQ
 Query Match 25.3%; Score 97.5; DB 22; Length 1445;
 Best Local Similarity 31.9%; Pred. No. 0.00086;
 Matches 22; Conservative 13; Mismatches 31; Indels 3; Gaps 2;
 QY 1 GDSFYIRVNLAMEGRAGKELQVHCNEVLHVDTMTFQG-CGCWHAHRVNSYTMKDTAAHGT 59
 DB 498 GDSFHIKTHFHCNPFSGEMAFKAGDVFRIIDTLHNGVVGSMQVLKIGRGHGE--MQRGV 555
 QY 60 IPNYSRAQQ 68
 DB 556 IPNKSRAEE 564
 RESULT 12
 AAU73251
 ID AAU73251 standard; Protein; 89 AA.
 XX
 AC AAU73251;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Mouse tight junction domain ZO-1, SH3 domain.
 XX
 KW Human; plakoglobin; cytostatic; osteopathic; dermatological; cardiant;
 KW plakoglobin related disease; skin carcinoma; acantholytic disease;
 KW basal cell carcinoma; squamous cell carcinoma; Naxos disease;
 KW extramammary Paget's disease; heart disease; skin blistering;
 KW subcorneal acantholysis; Grover's disease; Halley-Halley's disease;
 KW Darier's disease; ectodermal dysplasia; skin fragility syndrome.
 XX
 OS Mus sp.
 XX
 XX WO200185933-A2.
 XX
 PD 15-NOV-2001.
 XX

PF 02-MAY-2001; 2001WO-EP04872.
XX
XX
PR 09-MAY-2000; 2000EP-0201668.
XX
XX (VLAA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
PA
XX Van Roy F, Bonne S, Vanlandschoot A;
PI
XX WPI; 2002-062246/08.
DR
XX
XX New polypeptide, useful for treating skin carcinoma or acantholytic
PT disease such as Grover's and Darier's disease, comprises a protein
PT interacting with human plakoglobin and involved in transduction of
PT plakoglobin related signal to nucleus
XX
XX Example 4; Figure 4; 98pp; English.
PS
XX The invention relates to an isolated plakoglobin interacting polypeptide
CC (1). (1) is useful as a medicament and in the manufacture of a
CC medicament for treating plakoglobin related diseases, such as skin
CC carcinoma or an acantholytic disease, and to screen compounds that
CC interfere with the interaction of the polypeptide with plakoglobin
CC The plakoglobin related diseases include basal cell carcinoma, squamous
CC cell carcinoma, extramammary Paget's disease, Naxos disease, heart
CC diseases, skin blistering and acantholytic diseases such as subcorneal
CC acantholysis, Grover's disease, Hailey-Hailey's disease or Darier's
CC disease, and ectodermal dysplasia/skin fragility syndrome. AAU73245-
CC AAU73254 represent novel human plakoglobin interacting protein
CC amino acid sequences and related sequences of the invention.
XX
XX Sequence 89 AA;
SQ
Query Match 25.0%; Score 96.5; DB 23; Length 89;
Best Local Similarity 36.8%; Pred. No. 3.7e-05;
Matches 25; Conservative 9; Mismatches 31; Indels 3; Gaps 2;
QY 2 DSFYIRVNLAMEGRKALQVHCNEVLHVTDTMFQGCWHAHVNSYTMKDTAAHGTTI 60
Db 1 DSFYIRTHFEYKESPYGLSFNKGVEFRVDTLYNGKLGSLAIRIGNHKE--VERGII 58
QY 61 PNYSRAQQ 68
Db 59 PNKNRAEQ 66
RESULT 13
AAM93968
ID AAM93968 standard; Protein; 548 AA.
XX
XX AAM93968;
XX
XX 13-NOV-2001 (first entry)
DT
DE Human stomach cancer expressed polypeptide SEQ ID NO 4.
XX
XX Human; stomach cancer; marker; screening; micro-metastasis;
KW peritoneal dissemination.
XX Homo sapiens.
XX
XX WO200109317-A1.
PN
XX 08-FEB-2001.
PD
XX 28-JUL-2000; 2000WO-JP05063.
PF
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 18-OCT-1999; 99US-0119590.
PR 11-JAN-2000; 2000JP-0118776.
PR 17-FEB-2000; 2000US-0183322.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.
PA
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Aburatani H;
PI Kodama T, Migorikawa Y;
XX
XX WPI; 2001-570287/64.
DR N-PSDB; AAI93843.
DR
XX New Stomach cancer-associated genes, useful as markers in blood tests
PT for screening for the early stages of the disease
XX
XX Claim 1; Page 67-69; 242pp; Japanese.
PS
XX The invention relates to stomach cancer-expressed genes
CC (AAI93842-AAI93917) and the encoded proteins (AAM93967-AA94039). The
CC genes can be used as markers in blood tests for screening for the early
CC stages of the disease. The proteins and peptides can be used as targets
CC for screening for compounds to treat the disease. They can also be used
CC for predicting micro-metastases. The gene can predict peritoneal
CC dissemination.
XX
XX Sequence 548 AA;
SQ
Query Match 25.0%; Score 96.5; DB 22; Length 548;
Best Local Similarity 38.4%; Pred. No. 0.00035;
Matches 28; Conservative 7; Mismatches 33; Indels 5; Gaps 2;
QY 1 GDSFYIRVNLAMEGRKALQVHCNEVLHVTDTMFQGCWHHA--HRVNSYTMKD--TA 55
Db 104 GDSFYIRTHFELEPSPSPSGLGFTRGDFVHVLDTLHPGQSHARGGHWLAVRMGRLREQ 163
QY 56 AHGTIPNYSRAQQ 68
Db 164 ERGIIPNQSRAEQ 176
RESULT 14
AAB93753
ID AAB93753 standard; Protein; 548 AA.
XX
XX AAB93753;
XX
XX 26-JUN-2001 (first entry)
DT
DE Human protein sequence SEQ ID NO:13422.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW Homo sapiens.
XX
XX EPI074617-A2.
PN
XX 07-FEB-2001.
PD
XX 28-JUL-2000; 2000EP-0116126.
PF
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
PA
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
DR
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT

THIS PAGE BLANK (USPTO)


```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-562-737-21

Query Match          16.2%; Score 62.5; DB 4; Length 724;
Best Local Similarity 31.0%; Pred. No. 8.4;
Matches 22; Conservative 9; Mismatches 31; Indels 9; Gaps 3;

QY 4 FYIRVNLAME-----GRAGELQVHCNEVLHVDTMFGCGCWHAHVRVNSYTMKDTAAHG 58
||||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 431 FYIRALFDYDKTKDCGLLSQALSFFHFGDVLHVIDASDE--EWWQARRVHSDSETDDI--G 486

QY 59 TIPNYSRAQQQ 69
||: |::

Db 487 FIPSKRRVER 497

RESULT 3
US-09-562-737-27
; Sequence 27, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-562-737-27

Query Match          15.2%; Score 58.5; DB 4; Length 724;
Best Local Similarity 31.4%; Pred. No. 27;
Matches 22; Conservative 7; Mismatches 32; Indels 9; Gaps 3;

QY 4 FYIRVNLAME-----GRAGELQVHCNEVLHVDTMFGCGCWHAHVRVNSYTMKDTAAHG 58
||||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 431 FYIRALKDYDKTKDCGLLSQALSFFHFGDVLHVIDASNE--EWWQARRVQSDSETDDI--G 486

QY 59 TIPNYSRAQQ 68
||: |::

Db 487 FIPSKRRVER 496

RESULT 4
US-08-448-170-8
; Sequence 8, Application US/08448170
; Patent No. 5723758
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel
; APPLICANT: Cummings, David A.
; APPLICANT: Cannon, Raymond J.C.
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stelman, Steve
; TITLE OF INVENTION: No. 5723758el Bacillus thuringiensis Isolate Denoted
; TITLE OF INVENTION: B.t. PS158C2, Active Against Lepidopteran Pests, and Genes
; TITLE OF INVENTION: Encoding Lepidopteran-Active Toxins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
```

```
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,170
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/069,902
; FILING DATE: 01-JUNE-1993
; CLASSIFICATION: 424
; APPLICATION NUMBER: US 07/759,247
; FILING DATE: 13-SEPT-1991
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: M/S 102D.C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1227 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-448-170-8

Query Match          15.0%; Score 58; DB 1; Length 1227;
Best Local Similarity 29.9%; Pred. No. 59;
Matches 20; Conservative 12; Mismatches 31; Indels 4; Gaps 3;

QY 1 GDSFYIRVNLAMEGRAGELQVHCNEVLHVDTM-FQCGCWHAHVRVNSYTMKDTAAH-G 58
||||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1074 GRGYTLRYATYKEGYGEGCVTIH--EIENTDELAKFKNEEEVYPTDTGTCNDYTAHQG 1131

QY 59 TIPNYSR 65
||: |::

Db 1132 TAACNSR 1138

RESULT 5
US-08-961-803-9
; Sequence 9, Application US/08961803
; Patent No. 6150589
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel
; APPLICANT: Cummings, David A.
; APPLICANT: Cannon, Raymond J.C.
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stelman, Steve
; TITLE OF INVENTION: No. 6150589el Bacillus thuringiensis Isolate Denoted
; TITLE OF INVENTION: B.t. PS158C2, Active Against Lepidopteran Pests, and Genes
; TITLE OF INVENTION: Encoding Lepidopteran-Active Toxins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jay M. Sanders
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.803
; FILING DATE: 31-OCT-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/069,902
; FILING DATE: 01-JUNE-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,247
; FILING DATE: 13-SEPT-1991
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/448,170
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: M/S 102DCD1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (352) 375-8100
; TELEFAX: (352) 372-5800
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1227 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-961-803-9

Query Match 15.0%; Score 58; DB 4; Length 1227;
Best Local Similarity 29.9%; Pred. No. 59;
Matches 20; Conservative 12; Mismatches 31; Indels 4; Gaps 3;

Qy 1 GDSFYIRVNLAMEGRAGELQVHCNEVLHVTDTM-FQCGCGWHAHRVNSYTMKDTAAH-G 58
| : : ||| ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 1074 GRGYLRVTAYKEGYGECVTIH--E1ENNWDLKFKNCEEEVYPTDTGTCNDYTAHQG 1131
| : : ||| ||| : : : ||| : : : ||| : : : ||| : : : |||
Qy 59 TIPNYSR 65
| : : ||
Db 1132 TAACNSR 1138

RESULT 6
US-08-100-709-4
; Sequence 4, Application US/08100709
; Patent No. 5322687
; GENERAL INFORMATION:
; APPLICANT: Donovan, William P.
; APPLICANT: Tan, Yuping
; APPLICANT: Jany, Christine S.
; APPLICANT: Gonzalez Jr., Jose M.
; TITLE OF INVENTION: BACILLUS THURINGIENSIS cryET4 AND cryET5
; TITLE OF INVENTION: TOXIN GENES AND PROTEINS TOXIC TO LEPIDOPTERAN INSECTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S.
; ADDRESSEE: Nadel
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/176,865
; FILING DATE: 30-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/100,709
; FILING DATE: 29-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Egolf, Christopher
; REGISTRATION NUMBER: 27633
; REFERENCE/DOCKET NUMBER: 7205-49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-757-1590
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1229 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/100,709
; FILING DATE: 19930729
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Egolf, Christopher
; REGISTRATION NUMBER: 27633
; REFERENCE/DOCKET NUMBER: 7205-49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-757-1590
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1229 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-100-709-4

Query Match 15.0%; Score 58; DB 1; Length 1229;
Best Local Similarity 29.9%; Pred. No. 59;
Matches 20; Conservative 12; Mismatches 31; Indels 4; Gaps 3;

Qy 1 GDSFYIRVNLAMEGRAGELQVHCNEVLHVTDTM-FQCGCGWHAHRVNSYTMKDTAAH-G 58
| : : ||| ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 1076 GRGYLRVTAYKEGYGECVTIH--E1ENNWDLKFKNCEEEVYPTDTGTCNDYTAHQG 1133
| : : ||| ||| : : : ||| : : : ||| : : : ||| : : : |||
Qy 59 TIPNYSR 65
| : : ||
Db 1134 TAACNSR 1140

RESULT 7
US-08-176-865-4
; Sequence 4, Application US/08176865
; Patent No. 5616319
; GENERAL INFORMATION:
; APPLICANT: Donovan, William P.
; APPLICANT: Tan, Yuping
; APPLICANT: Jany, Christine S.
; APPLICANT: Gonzalez Jr., Jose M.
; TITLE OF INVENTION: BACILLUS THURINGIENSIS cryET4 AND cryET5
; TITLE OF INVENTION: TOXIN GENES AND PROTEINS TOXIC TO LEPIDOPTERAN INSECTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S.
; ADDRESSEE: Nadel
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/176,865
; FILING DATE: 30-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/100,709
; FILING DATE: 29-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Egolf, Christopher
; REGISTRATION NUMBER: 27633
; REFERENCE/DOCKET NUMBER: 7205-49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-757-1590
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1229 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; CURRENT APPLICATION DATA:
```

```
; MOLECULE TYPE: protein
US-08-176-865-4

Query Match          15.0%; Score 58; DB 1; Length 1229;
Best Local Similarity 29.9%; Pred. No. 59;
Matches 20; Conservative 12; Mismatches 31; Indels 4; Gaps 3;

QY 1 GDSFYIRVNLAMEGRAKELQVHCNEVLHVTDTM-FQCGCGWHAHRVNSYTMKDTAAH-G 58
   | : : ||| ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 1076 GRGYILRVTAKEGYGECVTIH--EIEENTDELKFKNCEEEVYPTDGTGCDNYTAHQG 1133

QY 59 TIPNYSR 65
   | : : |||
Db 1134 TAACNSR 1140

RESULT 8
US-08-474-038-4
; Sequence 4, Application US/08474038
; Patent No. 5679343
; GENERAL INFORMATION:
; APPLICANT: Donovan, William P.
; APPLICANT: Tan, Yuping
; APPLICANT: Jany, Christine S.
; APPLICANT: Gonzalez Jr., Jose M.
; TITLE OF INVENTION: BACILLUS THURINGIENSIS CRYET4 AND CRYET5
; TITLE OF INVENTION: TOXIN GENES AND PROTEINS TOXIC TO LEPIDOPTERAN INSECTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S.
; ADDRESSEE: Nadel
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,038
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,865
; FILING DATE: 30-DEC-1993
; APPLICATION NUMBER: US 08/100,709
; FILING DATE: 29-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Egolf, Christopher
; REGISTRATION NUMBER: 27633
; REFERENCE/DOCKET NUMBER: 7205-49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-757-1590
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1229 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-474-038-4

Query Match          15.0%; Score 58; DB 1; Length 1229;
Best Local Similarity 29.9%; Pred. No. 59;
Matches 20; Conservative 12; Mismatches 31; Indels 4; Gaps 3;

QY 1 GDSFYIRVNLAMEGRAKELQVHCNEVLHVTDTM-FQCGCGWHAHRVNSYTMKDTAAH-G 58
   | : : ||| ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 1076 GRGYILRVTAKEGYGECVTIH--EIEENTDELKFKNCEEEVYPTDGTGCDNYTAHQG 1133

QY 59 TIPNYSR 65
   | : : |||
Db 1134 TAACNSR 1140

RESULT 9
US-08-779-046-4
; Sequence 4, Application US/08779046
; Patent No. 5854053
; GENERAL INFORMATION:
; APPLICANT: Donovan, William P.
; APPLICANT: Tan, Yuping
; APPLICANT: Jany, Christine S.
; APPLICANT: Gonzalez Jr., Jose M.
; TITLE OF INVENTION: BACILLUS THURINGIENSIS CRYET4 AND CRYET5
; TITLE OF INVENTION: TOXIN GENES AND PROTEINS TOXIC TO LEPIDOPTERAN INSECTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S.
; ADDRESSEE: Nadel
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/779,046
; FILING DATE: 06-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/100,709
; FILING DATE: 29-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Egolf, Christopher
; REGISTRATION NUMBER: 27633
; REFERENCE/DOCKET NUMBER: 7205-49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-757-1590
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1229 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-779-046-4

Query Match          15.0%; Score 58; DB 2; Length 1229;
Best Local Similarity 29.9%; Pred. No. 59;
Matches 20; Conservative 12; Mismatches 31; Indels 4; Gaps 3;

QY 1 GDSFYIRVNLAMEGRAKELQVHCNEVLHVTDTM-FQCGCGWHAHRVNSYTMKDTAAH-G 58
   | : : ||| ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 1076 GRGYILRVTAKEGYGECVTIH--EIEENTDELKFKNCEEEVYPTDGTGCDNYTAHQG 1133

QY 59 TIPNYSR 65
   | : : |||
Db 1134 TAACNSR 1140

RESULT 10
US-08-881-340-4
; Sequence 4, Application US/08881340
; Patent No. 5942658
; GENERAL INFORMATION:
; APPLICANT: Donovan, William P.
; APPLICANT: Tan, Yuping
; APPLICANT: Jany, Christine S.
; APPLICANT: Gonzalez Jr., Jose M.
; TITLE OF INVENTION: BACILLUS THURINGIENSIS CRYET4 AND CRYET5
```

```
; TITLE OF INVENTION: TOXIN GENES AND PROTEINS TOXIC TO LEPTIDOPTERAN INSECTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S.
; ADDRESSEE: Nadel
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/881,340
; FILING DATE: 24-JUN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/100,709
; FILING DATE: 29-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Egolf, Christopher
; REGISTRATION NUMBER: 27633
; REFERENCE/DOCKET NUMBER: 7205-49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-757-1590
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1229 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-881-340-4

Query Match 15.0%; Score 58; DB 2; Length 1229;
Best Local Similarity 29.9%; Pred. No. 59;
Matches 20; Conservative 12; Mismatches 31; Indels 4; Gaps 3;

Qy 1 GDSFVIRVNLAMEGKAGELQVHCNEVLHVTDTM-FQCGCGWHHRVNSYTMKDTAAH-G 58
Db 1076 GRGYILRVYATKEGYGECVTH--EENNTDELKFKNCEEEVPTDTGTCDNYTAHQG 1133

Qy 59 TIPNYSR 65
Db 1134 TAACNSR 1140

RESULT 11
US-09-023-731-4
; Sequence 4, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-023-731-1
```

```
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 57
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-023-731-4

Query Match 14.9%; Score 57.5; DB 4; Length 57;
Best Local Similarity 32.6%; Pred. No. 1.8;
Matches 14; Conservative 9; Mismatches 15; Indels 5; Gaps 2;

Qy 15 RAKGELQVHCNEVLHVTDTMFGCG---CWHHRV--NSYTMK 52
Db 15 RSQGLTSSQHEIVHFTDVFIAAGSGPISCTVARHIIDNTSTTK 57

RESULT 12
US-09-023-731-1
; Sequence 1, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-023-731-1
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
;

```

```

; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; STRAIN: aizawai
; INDIVIDUAL ISOLATE: PS81A2
; IMMEDIATE SOURCE:
; LIBRARY: LambdaGem - 11 (tm) Library of August Sick
; CLONE: 81A2
US-08-040-751-3

Query Match      14.5%; Score 56; DB 1; Length 1174;
Best Local Similarity 29.5%; Pred. NO. 1e+02;
Matches 18; Conservative 12; Mismatches 27; Indels 4; Gaps 3;

QY 1 GDSFYIRVNLAMEGRAGKGELOVHCNEVLHVTDTM-FQGGCWHHRVNSYTMKDTAAH-G 58
DB 1017 GRGYILRVTAAYKEGYGECVTH--E1ENNDELAKFNCEEEVPTDTGTCNDYTAHOG 1074

QY 59 T 59
DB 1075 T 1075
```

Search completed: January 22, 2003, 08:57:39
Job time : 6.79839 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 2.82258 Seconds
(without alignments)
500.428 Million cell updates/sec

Title: US-09-767-215-2_COPY_676_745

Perfect score: 386

Sequence: 1 GDSFYIRVNLAMEGRAKEL.....MKDTAAHGTTIPNYSRAQQQL 70

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	386	100.0	1004	10	US-09-767-215-2
2	386	100.0	1138	10	US-09-767-215-5
3	145	37.6	1247	9	US-10-032-159A-8
4	98.5	25.5	1736	10	US-09-919-497-98
5	63	16.3	156	9	US-09-785-548-4
6	57.5	14.9	57	10	US-09-911-969-4
7	57.5	14.9	566	10	US-09-911-969-1
8	57	14.8	1050	10	US-09-866-562-38
9	56.5	14.6	1422	8	US-08-424-550B-82
10	54.5	14.1	767	10	US-09-919-497-59
11	54.5	14.1	4999	9	US-09-976-059-14
12	53.5	13.9	4999	9	US-09-976-059-15
13	52.5	13.6	510	10	US-09-742-684-4
14	52	13.5	121	10	US-09-764-877-1641
15	52	13.5	409	10	US-09-925-302-696
16	52	13.5	529	9	US-10-076-840-6
17	52	13.5	529	10	US-09-965-313-4
18	52	13.5	529	10	US-09-732-234-6
19	52	13.5	529	10	US-09-784-859-6

ALIGNMENTS

RESULT 1
US-09-767-215-2
; Sequence 2, Application US/09767215
; Patent No. US20020081636A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; FILE REFERENCE: 07334-142001
; CURRENT APPLICATION NUMBER: US/09/767,215
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/181,159
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1004
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-767-215-2

Query Match	100.0%	Score	386;	DB	10;	Length	1004;
Best Local Similarity	100.0%	Pred. No.	3.5e-41;				
Matches	70;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	GDSFYIRVNLAMEGRAKELQVHCNEVLHVTDTMFQCGCWHAHVRVNSYTMKDTHAAGTI	60				
Db	676	GDSFYIRVNLAMEGRAKELQVHCNEVLHVTDTMFQCGCWHAHVRVNSYTMKDTHAAGTI	735				
QY	61	PNYSRAQQQL 70					
Db	736	PNYSRAQQQL 745					
RESULT 2							
US-09-767-215-5							
; Sequence 5, Application US/09767215							
; Patent No. US20020081636A1							
; GENERAL INFORMATION:							
; APPLICANT: Bertin, John							
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED							
; PROTEIN FAMILY AND USES THEREOF							

; FILE REFERENCE: 07334-142001
; CURRENT APPLICATION NUMBER: US/09/767,215
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/181,159
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1138
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-767-215-5

Query Match 100.0%; Score 386; DB 10; Length 1138;
Best Local Similarity 100.0%; Pred. No. 4.1e-41;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDSFYIRVNLAMEGRAKGLQVHCNEVLHVDTMTFQGGCWHHRVNSYTMKDTAAHGTI 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 725 GDSFYIRVNLAMEGRAKGLQVHCNEVLHVDTMTFQGGCWHHRVNSYTMKDTAAHGTI 784

QY 61 PNYSAQQQL 70
|||||:|||||
Db 785 PNYSAQQQL 794

RESULT 3
US-10-032-159A-8
; Sequence 8, Application US/10032159A
; Patent No. US20020164703A1
; GENERAL INFORMATION:
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,
; FILE OF INVENTION: ENCODING NUCLEIC ACIDS, AND METHODS OF USE
; FILE REFERENCE: P-LJ 5100
; CURRENT APPLICATION NUMBER: US/10/032,159A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,457
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-159A-8

Query Match 37.6%; Score 145; DB 9; Length 1247;
Best Local Similarity 46.5%; Pred. No. 2.5e-10;
Matches 33; Conservative 12; Mismatches 24; Indels 2; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAKGLQVHCNEVLHVDTMTFQGGCWHHRVNSYTMKDTAAHGT 59
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 898 GDSFYIRLNLISQDAMTSLKCDVVHVDTMTYQDRHWRICARVDFPTDHDLDLDM-GT 956

QY 60 IPNSRAQQQL 70
|||||:|||||
Db 957 IPNSRAQQQL 967

RESULT 4
US-09-919-497-98
; Sequence 98, Application US/09919497
; Patent No. US20020108662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735

; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 98
; LENGTH: 1736
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-919-497-98

Query Match 25.5%; Score 98.5; DB 10; Length 1736;
Best Local Similarity 36.2%; Pred. No. 0.00032;
Matches 25; Conservative 9; Mismatches 32; Indels 3; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAKGLQVHCNEVLHVDTMTFQGGCWHHRVNSYTMKDTAAHGT 59
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 504 GDSFYIRTHFEVERESPGLSKNGKEVFRVAVDTLYNGKLSWLAIRICKNHKE--VERGI 561

QY 60 IPNSRAQQ 68
|||||:|||||
Db 562 IPKNKRAEQ 570

RESULT 5
US-09-785-548-4
; Sequence 4, Application US/09785548
; Patent No. US20020155577A1
; GENERAL INFORMATION:
; APPLICANT: AVENTIS PHARMACEUTICALS, INC.
; TITLE OF INVENTION: COMPOSITIONS THAT CAN BE USED FOR REGULATING THE ACTIVITY OF P
; FILE REFERENCE: ST00005
; CURRENT APPLICATION NUMBER: US/09/785,548
; CURRENT FILING DATE: 2001-02-20
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-785-548-4

Query Match 16.3%; Score 63; DB 9; Length 156;
Best Local Similarity 32.5%; Pred. No. 0.62;
Matches 13; Conservative 7; Mismatches 20; Indels 0; Gaps 0;

QY 2 DGSFYIRVNLAMEGRAKGLQVHCNEVLHVDTMTFQGGCWH 41
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 10 NSFVYVCKGPCQRCQVQPKLRVQCSTCRQATLTLTQGPSCW 49

RESULT 6
US-09-911-969-4
; Sequence 4, Application US/09911969
; Patent No. US20020137896A1
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; Akaihiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/911,969


```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/424.550B
; FILING DATE:
; CLASSIFICATION: 435435
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-424-550B-82

Query Match 14.6%; Score 56.5; DB 8; Length 1422;
Best Local Similarity 31.9%; Pred. No. 57;
Matches 15; Conservative 4; Mismatches 27; Indels 1; Gaps 1;

QY 23 HCNEVLHVTDTMFQCGCCGWAHRAHNSVTMKTDAAGHTIPNYSRAQQ 68
Db 202 YVOHWHVDPRTMFPCLXCNHLXRLMPCYRSNRHVRGHGKPNRSSIQK 248

RESULT 10
; Sequence 59, Application US/09919497
; Patent No. US20020106662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 59
; LENGTH: 767
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-919-497-59

Query Match 14.1%; Score 54.5; DB 10; Length 767;
Best Local Similarity 29.6%; Pred. No. 50;
Matches 21; Conservative 9; Mismatches 32; Indels 9; Gaps 3;

QY 4 FYIRVNLAME-----GRAGELQVHCNEVLHVTDTMFQCGCGWAHRAHNSVTMKTDAAGH 58
Db 474 FYIRALFDYDKTKDCGFLSQALSFRFGDVLHVLDASDE---ENWQARRVHSDSETDDI--G 529
QY 59 TIPNYSRAQQ 69
Db 530 TIPSKRRVERR 540

RESULT 11
; Sequence 14, Application US/09976059
; Patent No. US20020164747A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
```

```
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin
; FILE REFERENCE: 3019-PCT
; CURRENT APPLICATION NUMBER: US/09/976.059
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 4999
; TYPE: PRT
; ORGANISM: Actinoplanes sp.
; US-09-976-059-14

Query Match 14.1%; Score 54.5; DB 9; Length 4999;
Best Local Similarity 30.0%; Pred. No. 4.6e+02;
Matches 18; Conservative 3; Mismatches 22; Indels 17; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAGELQVHCNEVLHVTDTMFQCGCGW-----HAHRVNSY 49
Db 4846 GAAVYRMYTSGTGRPKGVVTHQNLDLATDT-----CWGPTPRVLHFHAPHAFDASSY 4899

RESULT 12
; Sequence 15, Application US/09976059
; Patent No. US20020164747A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin
; FILE REFERENCE: 3019-PCT
; CURRENT APPLICATION NUMBER: US/09/976.059
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 4999
; TYPE: PRT
; ORGANISM: Actinoplanes sp.
; US-09-976-059-15

Query Match 13.9%; Score 53.5; DB 9; Length 4999;
Best Local Similarity 30.0%; Pred. No. 6.2e+02;
Matches 18; Conservative 2; Mismatches 23; Indels 17; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAGELQVHCNEVLHVTDTMFQCGCGW-----HAHRVNSY 49
Db 3789 GAEVYRMYTSGTGRPKGVVTHQNLDLATDT-----CWGPTPRVLHFHAPHAFDASSY 3842

RESULT 13
; Sequence 4, Application US/09742684
; Patent No. US2001003036A1
; GENERAL INFORMATION:
; APPLICANT: Mathews, Lawrence S.
; FILE REFERENCE: Tsuchida, Kunihiko
; TITLE OF INVENTION: CLONING AND RECOMBINANT PRODUCTION OF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
;
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/742,684
; FILING DATE: 19-Dec-2000
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,123
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/300,584
; FILING DATE: 02-SEP-1994
; APPLICATION NUMBER: US 07/880,220
; FILING DATE: 08-MAY-1992
; APPLICATION NUMBER: US 07/773,229
; FILING DATE: 09-OCT-1991
; APPLICATION NUMBER: US 07/698,709
; FILING DATE: 10-MAY-1991
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9927
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
;
; INFORMATION FOR SEQ ID NO: 4:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 510 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-742-684-4
Query Match      13.6%; Score 52.5; DB 10; Length 510;
Best Local Similarity 46.2%; Pred. No. 55;
Matches 12; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

Qy 17 KGELOVHCNEVLHVTDTMFQCGGWH 42
Db 276 KGNL-VSNELCHITETMARGLAYLH 300

RESULT 14
US-09-764-877-1641
; Sequence 1641, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1641
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (10)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

;
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/742,684
; FILING DATE: 19-Dec-2000
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,123
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/300,584
; FILING DATE: 02-SEP-1994
; APPLICATION NUMBER: US 07/880,220
; FILING DATE: 08-MAY-1992
; APPLICATION NUMBER: US 07/773,229
; FILING DATE: 09-OCT-1991
; APPLICATION NUMBER: US 07/698,709
; FILING DATE: 10-MAY-1991
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9927
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
;
; INFORMATION FOR SEQ ID NO: 4:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 510 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-742-684-4
Query Match      13.6%; Score 52.5; DB 10; Length 510;
Best Local Similarity 46.2%; Pred. No. 55;
Matches 12; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

Qy 17 KGELOVHCNEVLHVTDTMFQCGGWH 42
Db 276 KGNL-VSNELCHITETMARGLAYLH 300

RESULT 15
US-09-925-302-696
; Sequence 696, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 696
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (111)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (208)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-696
Query Match      13.5%; Score 52; DB 10; Length 409;
Best Local Similarity 27.0%; Pred. No. 49;
Matches 17; Conservative 10; Mismatches 16; Indels 20; Gaps 4;

Qy 5 YIRVNLAEGRAKGELOVHCN-----EVLH--VTDTMFQCGG-----CWHHR 45
Db 28 YIRNIQAK-RMQGLEAHVNGFRFTSVRGDKVDILYNNIRKHALFPQCGDEMIIVLHFHL 86
Qy 46 VNS 48
Db 87 KNA 89

Search completed: January 22, 2003, 08:52:54
Job time : 3.82258 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 5.3629 Seconds
(without alignments)
1254.807 Million cell updates/sec

Title: US-09-767-215-2_COPY_676_745
Perfect score: 386
Sequence: 1 GDSFYIRVNLAMEGRAGKEL.....MKDTAAHGTTIPNYSRAQQQL 70
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105.5	27.3	1034	2 T22166	hypothetical prote
2	102.5	26.6	1745	2 A46431	tight junction-ass
3	101	26.2	775	2 I46236	tight junction pro
4	100	25.9	1116	2 I54378	gene x104 protein
5	100	25.9	1163	2 JE0366	tight junction pro
6	98.5	25.5	1736	2 A47747	tight junction pro
7	97.5	25.3	1367	2 T13703	tama protein - fru
8	78.5	20.3	1281	2 T00346	hypothetical prote
9	63.5	16.5	687	1 HYRTH	thimet oligopeptid
10	60	15.5	431	2 D70185	3-hydroxy-3-methyl
11	59.5	15.4	206	2 C36365	transforming prote
12	59.5	15.4	299	2 A13184	transcription regu
13	59.5	15.4	411	2 H84495	probable retroelem
14	59.5	15.4	864	2 H84176	bacterio-opsin act
15	59	15.3	227	2 F83505	hypothetical prote
16	59	15.3	415	2 T23215	hypothetical prote
17	58.5	15.2	541	2 A12238	glutamine-fructose
18	58.5	15.2	1164	2 G82100	DNA polymerase III
19	58	15.0	686	2 A55665	microtubule-assoc
20	57	14.8	105	2 E82386	hypothetical prote
21	57	14.8	161	2 T26727	hypothetical prote
22	57	14.8	372	2 A72175	DI4R protein - var
23	57	14.8	779	2 T32811	hypothetical prote
24	57	14.8	1050	2 JW0092	serine-threonine k
25	56.5	14.6	865	2 T15771	hypothetical prote
26	56	14.5	349	2 T36152	probable transposa
27	56	14.5	386	2 G81256	probable sulfate a
28	56	14.5	517	1 ERA047	early E2A DNA-bind
29	56	14.5	1096	2 T13802	centrosome-associa

ALIGNMENTS

RESULT 1
T22166
hypothetical protein F44D12.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22166
R;Coles, L.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19525
A;Accession: T22166
A;Status: preliminary; translated from GB/EMBL/DBBJ
A;Molecule type: DNA
A;Residues: 1-1034 <WIL>
A;Cross-references: EMBL:Z68298; PIDN:CAA92607.1; GSPDB:GN00022; CESP:F44D12.1
A;Experimental source: clone F44D12
C;Genetics:
A;Gene: CESP:F44D12.1
A;Map position: 4
A;Introns: 46/2; 105/3; 187/3; 283/1; 365/3; 444/2; 501/2; 605/1; 688/1; 922/2; 980/3

Query Match 27.3%; Score 105.5; DB 2; Length 1034;
Best Local Similarity 35.2%; Pred. No. 9.1e-05;
Matches 25; Conservative 12; Mismatches 31; Indels 3; Gaps 2;

Qy 1 GDSFYIRVNLAMEGRAGKELQVHCNEVLHVTDTFQG-CGCWHAHRVNSYTMKDTAAHGT 59
||||:||||: |||:|:|:| |||:|:|:| |||:|:|:| |||:|:|:| |||:|:|:|
Db 753 GDSFLLRVNIDRSMENKDELDKCGDVVVDKTLMLMGTGRWAKVKDEGRQ--REHGA 810

Qy 60 IPNYSRAQQQL 70
||:|:|:|
Db 811 IPSSTTVVQAI 821

RESULT 2
A46431
tight junction-associated protein ZO-1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: A46431
R;Ittoh, M.; Nagafuchi, A.; Yonemura, S.; Kitani-Yasuda, T.; Tsukita, S.;
J. Cell Biol. 121, 491-502, 1993
A;Title: The 220-kD protein colocalizing with cadherins in non-epithelial cells is id
py.
A;Reference number: A46431; MUID:93252986; PMID:8486731
A;Accession: A46431
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-1745 <ITO>
A;Cross-references: GB:D14340; NID:g303709; PIDN:BAA03274.1; PID:d1003784; PID:g30371
A;Experimental source: F9 cells
A;Note: sequence extracted from NCBI backbone (NCBIN:131200, NCBIIP:131201)
C;Superfamily: guanylate kinase homology; GLCF domain homology

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 2.72849 Seconds
(without alignments)
1064.082 Million cell updates/sec

Title: US-09-767-215-2_COPY_676_745

Perfect score: 386

Sequence: 1 GDSFYRVNLAWEGRAGEL.....MKDTAAHGTFIPNYSRAQQQL 70

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	386	100.0	1004	1 CARE_HUMAN	Q9bx16 homo sapien
2	283.5	73.4	999	1 CARE_MOUSE	Q9Kf0 mus musculus
3	146	37.8	1147	1 CARE_HUMAN	Q9bx17 homo sapien
4	116	30.1	1032	1 CARE_HUMAN	Q9bwt7 homo sapien
5	106	27.5	1021	1 CARE_MOUSE	P58660 mus musculus
6	105	27.2	905	1 Z03_MOUSE	Q9qxy1 mus musculus
7	102.5	26.6	1745	1 Z01_MOUSE	P39447 mus musculus
8	101	26.2	1174	1 Z02_CANFA	Q95168 canis famill
9	100	25.9	1190	1 Z02_HUMAN	Q9udy2 homo sapien
10	99	25.6	1167	1 Z02_MOUSE	Q9z0u1 mus musculus
11	98.5	25.5	1736	1 Z01_HUMAN	Q07157 homo sapien
12	96.5	25.0	933	1 Z03_HUMAN	Q95049 homo sapien
13	94.5	24.5	898	1 Z03_CANFA	O62583 canis famill
14	63.5	16.5	686	1 MEFD_RAT	P24155 rattus norv
15	62.5	16.2	664	1 EMAP_LYTV	Q9yic1 lytechinus
16	62.5	16.2	724	1 DLG4_MOUSE	O62108 mus musculus
17	60	15.5	431	1 HMDH_BORBU	O51628 borrelia bu
18	59.5	15.4	205	1 RAS3_RHRA	P22280 rhizomucor
19	59	15.3	415	1 BCAT_CAEEL	P54688 caenorhabdi
20	58.5	15.2	1159	1 DP3A_VIBCH	P52022 vibrio chol
21	58	15.0	686	1 EMAP_STRPU	Q26613 strongyloce
22	58	15.0	1227	1 C1BB_BACTU	O85805 bacillus th
23	58	15.0	1229	1 C1BB_BACTU	Q45739 bacillus th
24	58	15.0	1231	1 C1BD_BACT2	Q9za25 bacillus th
25	57	14.8	1050	1 BU1B_HUMAN	O60566 homo sapien
26	56	14.5	517	1 DNE2_ADE07	P04497 human adeno
27	56	14.5	1174	1 C1EB_BACTA	Q03745 bacillus th
28	55.5	14.4	344	1 HYGH_HYONI	P24397 hyocysamus
29	55.5	14.4	368	1 VGH3_HCMVA	P08560 human cytom
30	55	14.2	513	1 COT4_BACSU	P07788 bacillus su
31	55	14.2	531	1 NXF3_HUMAN	Q9H4d5 homo sapien
32	55	14.2	766	1 METE_PSEAE	P57703 pseudomonas
33	55	14.2	1683	1 POLG_DEN2T	P27914 dengue viru

RESULT 1

ID	CARE_HUMAN	STANDARD;	PRT;	1004 AA.
AC	Q9BXL6; O9BVB5;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Caspase recruitment domain protein 14 (CARD-containing MAGUK protein			
DE	2) (Carma 2).			
GN	CARD14 OR CARMA2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-21192234; PubMed-11278692;			
RA	Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,			
RA	Srinivasula S.M., Merriam S., DiStefano P.S., Alnemri E.S.;			
RT	"CARD11 and CARD14 are novel caspase recruitment domain			
RT	(CARD)/membrane-associated guanylate kinase (MAGUK) family members			
RT	that interact with Bcl10 and activate NF-kappaB.";			
RL	J. Biol. Chem. 276:11877-11882(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-21255663; PubMed-11356195;			
RA	Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;			
RT	"Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10			
RT	phosphorylation and NF-kappaB activation.";			
RL	FEBS Lett. 496:121-127(2001).			
RN	[3]			
RP	ERRATUM.			
RA	Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;			
RL	FEBS Lett. 505:198-198(2001).			
RN	[4]			
RP	SEQUENCE OF 1-740 FROM N.A.			
RC	TISSUE=Cervix, and Colon;			
RA	Strausberg R.;			
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the			
CC	phosphorylation of Bcl10.			
CC	-!- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD			
CC	interaction.			
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic.			
CC	-!- TISSUE SPECIFICITY: Expressed in placenta. Also detected in HeLa			
CC	S3 cells, but not in the other cancer cell lines tested.			
CC	-!- SIMILARITY: CONTAINS 1 CARD DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.			
CC	-!- CAUTION: Supposed to contain a SH3 domain which is not detected by			
CC	PROSITE, Pfam or SMART.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			

ALIGNMENTS

34	55	14.2	3388	1	POLG_DEN2P	P12823 d genome po
35	55	14.2	3391	1	POLG_DEN2J	P07564 d genome po
36	55	14.2	3391	1	POLG_DEN2N	P14340 d genome po
37	54.5	14.1	673	1	REP_ECOLI	P09980 escherichia
38	54.5	14.1	674	1	REP_SALTY	Q91681 salmonella
39	54.5	14.1	767	1	DLG4_HUMAN	P78352 homo sapien
40	54	14.0	3391	1	POLG_DEN26	P29590 d genome po
41	54	14.0	3391	1	POLG_DEN27	P29591 d genome po
42	53.5	13.9	115	1	YV1_TYLCM	P27270 tomato yell
43	53.5	13.9	123	1	RS13_MVCBO	P45813 mycobacteri
44	53.5	13.9	243	1	Y762_PYRHO	O58498 pyrococcus
45	53.5	13.9	563	1	IDS_MOUSE	Q08890 mus musculus


```
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -!- CAUTION: Supposed to contain a SH3, a PDZ and a guanylate kinase-
CC like domain. But none of these 3 domains are detected by PROSITE,
CC Pfam or SMART.
CC
CC -!- CAUTION: Ref.4 sequence differs from that shown due to various
CC gene identification problems.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AY028896; AAK26165.1; -
CC EMBL; AY032928; AAK54454.1; -
CC EMBL; AL049851; CAB63075.1; ALT_SEQ.
CC EMBL; AL049851; CAB63076.1; ALT_SEQ.
CC EMBL; AL022315; CAB42832.1; ALT_SEQ.
CC PROSITE; PS50209; CARD; 1.
CC Coiled coil. 23 115 CARD.
CC DOMAIN 138 456 COILED COIL (POTENTIAL).
CC DOMAIN 567 574 POLY-SER.
CC CONFLICT 289 289 Q -> R (IN REF. 4).
CC CONFLICT 917 917 K -> KO (IN REF. 4; CAB63075).
CC CONFLICT 932 932 R -> L (IN REF. 4).
CC SEQUENCE 1032 AA; 115946 MW; 8377319AB82A0949 CRC64;
SQ
Query Match 30.1%; Score 116; DB 1; Length 1032;
Best Local Similarity 40.0%; Pred. No. 1.6e-06;
Matches 28; Conservative 11; Mismatches 29; Indels 2; Gaps 2;
QY 2 DSFYIRVNLAMEGRAGE-LQVHCNEVLHVDTMFGQCGCWHAHVRVSYTMKDTAAHGTTI 60
Db 705 EPFYIRANLTLPERDHALCVKAKQEILRLVDSAYKRRQEWFCFTRVDPLTLRD-LDRGTV 763
QY 61 PNYSQAOQLL 70
Db 764 PNYQRAOQLL 773
RESULT 5
CARA_MOUSE
ID CARA_MOUSE STANDARD; PRT; 1021 AA.
AC P58660.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Caspase recruitment domain protein 10 (Bcl10-interacting MAGUK protein
DE 1) (Bim1).
DE CARD10 OR BIM1.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21391892; PubMed=11387339;
RA McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A.,
RA Li Q., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W.,
RA Nunez G.;
RT "Bim1, a MAGUK family member linking protein kinase c activation to
RT Bcl10-mediated NF-kappa B induction.",
RL J. Biol. Chem. 276:30589-30597(2001).
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC -!- SUBUNIT: CARD10 and Bcl10 bind to each other by CARD-CARD
CC interaction. They both participate in a complex with MALT1, where
CC MALT1 binds to Bcl10.
CC -!- TISSUE SPECIFICITY: Highly expressed in kidney, heart followed by
CC brain, lung, liver, skeletal muscle and testis.
CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
```

```
CC -!- CAUTION: Supposed to contain a SH3, a PDZ and a guanylate kinase-
CC like domain. But none of these 3 domains are detected by PROSITE,
CC Pfam or SMART.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF363456; AAK60136.1; -
CC PROSITE; PS50209; CARD; 1.
CC Coiled coil. 23 115 CARD.
CC DOMAIN 138 450 COILED COIL (POTENTIAL).
CC DOMAIN 558 565 POLY-SER.
CC SEQUENCE 1021 AA; 114413 MW; 4811A09BDB8F792C CRC64;
SQ
Query Match 27.5%; Score 106; DB 1; Length 1021;
Best Local Similarity 37.1%; Pred. No. 2.7e-05;
Matches 26; Conservative 12; Mismatches 30; Indels 2; Gaps 2;
QY 2 DSFYIRVNLAMEGRAGE-LQVHCNEVLHVDTMFGQCGCWHAHVRVSYTMKDTAAHGTTI 60
Db 696 EPFYIRANFSLPERSDHALCVKAKQEILRLVDPAHKRRQEWFCFTRVDPLTLRD-LDRGTV 754
QY 61 PNYSQAOQLL 70
Db 755 PNYQRAOQLL 764
RESULT 6
Z03_MOUSE
ID Z03_MOUSE STANDARD; PRT; 905 AA.
AC Q90XY1.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona
DE occludens 3 protein) (Tight junction protein 3).
GN TJP3 OR Z03.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BALB/c;
RX MEDLINE=20069797; PubMed=10601346;
RA Itoh M., Furuse M., Morita K., Kubota K., Saitou M., Tsukita S.;
RT "Direct binding of three tight junction-associated MAGUKs, ZO-1, ZO-2,
RT and ZO-3, with the COOH termini of claudins.",
RL J. Cell Biol. 147:1351-1363(1999).
CC -!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-1.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF157006; AAF24175.1; -
CC HSP; P31016; 1BFE.
CC MGD; MGI:1351650; Tjp3.
CC InterPro; IPR000619; Guanylate_kin.
```

```

DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS00002; SH3; FALSE_NEG.
DR PROSITE; PS0106; PDZ; 3.
KW Tight junction; SH3 domain; Repeat; Membrane.
FT DOMAIN 11 93 PDZ 1.
FT DOMAIN 187 264 PDZ 2.
FT DOMAIN 368 434 PDZ 3.
FT DOMAIN 467 540 SH3.
FT DOMAIN 654 754 GUANYLATE_KINASE.
SQ SEQUENCE 905 AA; 99324 MW; B787BA1592661FEE CRC64;

Query Match 27.2%; Score 105; DB 1; Length 905;
Best Local Similarity 37.2%; Pred. No. 3.2e-05;
Matches 29; Conservative 8; Mismatches 29; Indels 12; Gaps 2;

QY 1 GDSFYRVNLAMEGRAKGELOVHCNEVLHVDTMFGCGCWHHRVNSYTMKDTAA---- 56
DB 464 GDSFYRTHFELEPPYGLGTRGDVHVDTLXPQSGPHGH--SSHGGLWLAARMGR 521

QY 57 -----HGTPNVSRAQQ 68
DB 522 DLREQGVIPNQSRREQ 539

RESULT 7
ZOL_MOUSE STANDARD; PRT; 1745 AA.
AC P39447;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-1 (zonula occludens 1 protein) (Zona
DE occludens 1 protein) (tight junction protein 1).
GN TJP1 OR ZOL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=93252986; PubMed=8486731;
RA Itoh M., Nagafuchi A., Yonemura S., Yasuda-Kitani T., Tsukita S.,
RA Tsukita S.;
RT "The 220-kD protein colocalizing with cadherins in non-epithelial
RT cells is identical to ZO-1, a tight junction-associated protein in
RT epithelial cells: cDNA cloning and immunoelectron microscopy.";
RL J. Cell Biol. 121:491-502(1993).
CC -!- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL
CC REQUIRED FOR TIGHT JUNCTION ASSEMBLY. WHILE THE C-TERMINAL MAY
CC HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN
CC MIGHT BE INVOLVED IN STABILIZING JUNCTIONS.
CC -!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-3.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.
CC MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT
CC OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE_KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14340; BAA03274.1; -.
DR PIR; A46431; A46431.
DR HSP; P31016; IBE9.
DR MGD; MGI:98759; Tjp1.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00791; ZU5; 1.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00218; ZU5; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
DR PROSITE; PS00002; SH3; 1.
KW Tight junction; SH3 domain; Repeat; Membrane.
FT DOMAIN 23 110 PDZ 1.
FT DOMAIN 186 264 PDZ 2.
FT DOMAIN 421 502 PDZ 3.
FT DOMAIN 516 584 SH3.
FT DOMAIN 644 794 GUANYLATE_KINASE.
FT DOMAIN 1242 1247 POLY-PRO.
FT DOMAIN 1424 1430 POLY-PRO.
SQ SEQUENCE 1745 AA; 194710 MW; C3DA2C0A9F411F66 CRC64;

Query Match 26.6%; Score 102.5; DB 1; Length 1745;
Best Local Similarity 37.7%; Pred. No. 0.00013;
Matches 26; Conservative 9; Mismatches 31; Indels 3; Gaps 2;

QY 1 GDSFYRVNLAMEGRAKGELOVHCNEVLHVDTMFGG-CGCHHHRVNSYTMKDTAAHGT 59
DB 516 GDSFYRTHFEYEKESPYGLSPKGEVFRVDTLYNGKLGSLAIRGNHKE--VERGI 573

QY 60 IPNYSRAQQ 68
DB 574 IPKNRAEQ 582

RESULT 8
ZOL_MOUSE STANDARD; PRT; 1174 AA.
AC Q95168;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-2 (zonula occludens 2 protein) (Zona
DE occludens 2 protein) (tight junction protein 2).
GN TJP2 OR ZO2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96421547; PubMed=8824195;
RA Beatch M., Jesaitis L.A., Gallin W., Goodenough D.A., Stevenson B.R.;
RT "The tight junction protein ZO-2 contains three PDZ (PSD-95/Discs-
RT Large/ZO-1) domains and an alternatively spliced region.";
RL J. Biol. Chem. 271:25723-25726(1996).
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94179414; PubMed=8132716;
RA Jesaitis L.A., Goodenough D.A.;
RT "Molecular characterization and tissue distribution of ZO-2, a tight

```



```

QY 60 IPNYSRAQQ 68
DB 562 IPKNRAEQ 570

RESULT 12
Z03_HUMAN
ID Z03_HUMAN STANDARD; PRT; 933 AA.
AC O95049;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona
DE occludens 3 protein) (tight junction protein 3).
GN TJP3 OR ZO3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Brower A., Gordon L., Dias J., Ramirez M.,
RA Stilwagen S., Phan H., Velasco N., Do L., Regata W., Terry A.,
RA Barnes J., Danganan L., Elier A., Christensen M., Georgescu A.,
RA Avila J., Liu S., Attix C., Andreise T., Trankheim M.,
RA Anico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R.,
RA Thomas P., Quan G., Krommiller B., Arellano A., Sanders C., Ow D.,
RA Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.;
RA "Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
RT serine protease gene cluster.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-1 (BY
CC SIMILARITY).
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC005954; AAC72274.1; ALT_INIT.
DR HSP; P31016; 1BFE.
DR Genew; HGNC:11829; TJP3.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50106; PDZ; 3.
KW Tight junction; SH3 domain; Repeat; Membrane.
FT DOMAIN 11 93
FT PDZ 1.
FT PDZ 2.
FT PDZ 3.
FT DOMAIN 195 272
FT PDZ 1.
FT PDZ 2.
FT PDZ 3.
FT DOMAIN 394 460
FT PDZ 1.
FT PDZ 2.
FT PDZ 3.
FT DOMAIN 489 563
FT PDZ 1.
FT PDZ 2.
FT PDZ 3.
FT DOMAIN 675 775
FT PDZ 1.
FT PDZ 2.
FT PDZ 3.
SQ SEQUENCE 933 AA; 102800 MW; 680298CFD0615B47 CRC64;

Query Match 25.0%; Score 96.5; DB 1; Length 933;
Best Local Similarity 38.4%; Pred. No. 0.00038;

Matches 27; Conservative 7; Mismatches 32; Indels 9; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAKELQVHCNEVLHVTDTMTFGCG--TA 55
DB 489 GDSFYIRTHFELEPPSPGCGTGRGDVFNHLDLHPGQSHARGHVLAVRMGRDLREQ 548
QY 56 AHGTIPNYSRAQQ 68
DB 549 ERGIIPNQSRAEQ 561

Matches 28; Conservative 7; Mismatches 33; Indels 5; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAKELQVHCNEVLHVTDTMTFGCGCGCWAH---HRVNSYTMKD--TA 55
DB 489 GDSFYIRTHFELEPPSPGCGTGRGDVFNHLDLHPGQSHARGHVLAVRMGRDLREQ 548
QY 56 AHGTIPNYSRAQQ 68
DB 549 ERGIIPNQSRAEQ 561

RESULT 13
Z03_CANFA
ID Z03_CANFA STANDARD; PRT; 898 AA.
AC O62683;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona
DE occludens 3 protein) (tight junction protein 3).
GN TJP3 OR ZO3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RA MEDLINE; 98198478; PubMed=9531559;
RA Haskins J., Gu L., Wittchen E.S., Hibbard J., Stevenson B.R.;
RT "ZO-3, a novel member of the MAGUK protein family found at the tight
RT junction, interacts with ZO-1 and occludin.";
RL J. Cell Biol. 141:199-208(1998).
CC -!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-1.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF023617; AAC39177.1; -.
DR HSP; P31016; 1BFE.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
KW Tight junction; SH3 domain; Repeat; Membrane.
FT DOMAIN 11 93
FT PDZ 1.
FT PDZ 2.
FT PDZ 3.
FT DOMAIN 187 264
FT PDZ 1.
FT PDZ 2.
FT PDZ 3.
FT DOMAIN 369 435
FT PDZ 1.
FT PDZ 2.
FT PDZ 3.
FT DOMAIN 464 538
FT PDZ 1.
FT PDZ 2.
FT PDZ 3.
FT DOMAIN 650 750
FT PDZ 1.
FT PDZ 2.
FT PDZ 3.
SQ SEQUENCE 898 AA; 98414 MW; 8091D6132DB9F15D CRC64;

Query Match 24.5%; Score 94.5; DB 1; Length 898;
Best Local Similarity 36.0%; Pred. No. 0.00064;
Matches 27; Conservative 7; Mismatches 32; Indels 9; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAKELQVHCNEVLHVTDTMTFGCG-----CWAHVRVNSYTMKD 53

```


THIS PAGE BLANK (USPTO)


```

Db 811 IPSSTTVQAI 821
      ||.: |.: |.:
      InterPro; IPR001478; PDZ.
      InterPro; IPR001452; SH3.
      InterPro; IPR00906; ZU5.
      Pfam; PF00625; Guanylate_kin; 1.
      Pfam; PF00595; PDZ; 3.
      Pfam; PF00018; SH3; 1.
      Pfam; PF00791; ZU5; 1.
      PRINTS; PR01597; ZONOCCLUDNS.
      PRINTS; PR01598; ZONOCCLUDNS1.
      PRINTS; PR01600; ZONOCCLUDNS3.
      SMART; SM00072; GuKc; 1.
      SMART; SM00228; PDZ; 3.
      SMART; SM00326; SH3; 1.
      SMART; SM00218; ZU5; 1.
      PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
      PROSITE; PS0106; PDZ; 3.
      PROSITE; PS50002; SH3; 1.
      KW SH3 domain.
      SQ SEQUENCE 1769 AA; 197606 MW; 181E9F36CEBC96EF CRC64;

Query Match 26.6%; Score 102.5; DB 6; Length 1769;
Best Local Similarity 37.7%; Pred. No. 0.00012;
Matches 26; Conservative 9; Mismatches 31; Indels 3; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAKGLQVHCNEVLHVTDTMTFOG-CGCWHAHRVNSYTKMDTAHGT 59
      ||||| |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.:
Db 515 GDSFYIRTHFEYKESPYGLSEFNKGEVFRVVDVTLYPGSGPGHGH--SSHGLWLAARMGR 520
      ||||| |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.:
QY 60 IPNYSRAQQ 68
      ||| :||:|
Db 573 IPKNRAEQ 581
      ||| :||:|

RESULT 7
P70625 PRELIMINARY; PRT; 813 AA.
AC P70625; P97625;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Zonula occludens 2 protein (Fragment).
GN ZO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR-KYOTO;
RX MEDLINE=20027749; PubMed=10559001;
RA Adams L.D., Lemire J.M., Schwartz S.M.;
RT "A systematic analysis of 40 random genes in cultured vascular smooth
RT muscle subtypes reveals a heterogeneity of gene expression and
RT identifies the tight junction gene zonula occludens 2 as a marker of
RT epithelioid 'pup' smooth muscle cells and a participant in carotid
RT neointimal formation."
RT Arterioscler. Thromb. Vasc. Biol. 19:2600-2608(1999).
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; U75916; ABA6979.1; -.
DR HSP; P31016; IBE9.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR01597; ZONOCCLUDNS.
DR PRINTS; PR01599; ZONOCCLUDNS2.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 1.

Query Match 27.2%; Score 105; DB 11; Length 904;
Best Local Similarity 37.2%; Pred. No. 2.4e-05;
Matches 29; Conservative 8; Mismatches 29; Indels 12; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAKGLQVHCNEVLHVTDTMTFOGCCWHAHRVNSYTKMDTAA--- 56
      ||||| |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.:
Db 463 GDSFYIRTHFELEPPYPGLGTRGDFVHVDTLYPGSGPGHGH--SSHGLWLAARMGR 520
      ||||| |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.: |.:
QY 57 -----HGIPNYSRAQQ 68
      | ||| |||:|
Db 521 DLREQERGVIPONSRAEQ 538
      | ||| |||:|

RESULT 6
O97758 PRELIMINARY; PRT; 1769 AA.
AC O97758;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ZO-1 MDCK.
GN ZO1-MDCK.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BREED; COCKER SPANIEL;
RX MEDLINE=99196918; PubMed=10094817;
RA Gonzalez-Mariscal L., Islas S., Contreras R.G., Garcia-Villegas M.R.,
RA Betanzos A., Vega J., Diaz-Quinonez A., Martin-Orozco N.,
RA Ortiz-Navarrete V., Cerejido M., Valdes J.;
RT "Molecular characterization of the tight junction protein ZO-1 in MDCK
RT cells."
RT Exp. Cell Res. 248:97-109(1999).
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; U55935; AAD11529.1; -.
DR HSP; P31016; IBE9.
DR InterPro; IPR000619; Guanylate_kin.

```



```

DR FlyBase; FBgn00031177; pyd.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
DR SEQUENCE 1367 AA; 148347 MW; 571C4566C6B68BF8 CRC64;

Query Match      25.3%; Score 97.5; DB 5; Length 1367;
Best Local Similarity 31.9%; Pred. No. 0.00042;
Matches 22; Conservative 13; Mismatches 31; Indels 3; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAKELQVHCNEVLHVTDTMQG-CGCWHAHRVNSYTMKDTAAHGT 59
||||: : : :|||: :| |||: |||: :|
DB 498 GDSFHKTHFCHDNPCKGEMAFKAGDVRVIDTLHNGVVGSMQVLKIGRGHQE--MQRGV 555

QY 60 IPNYSRAQ 68
||| |||::
DB 556 IPNKSRAEE 564

RESULT 11
Q9VHK4 PRELIMINARY; PRT; 1445 AA.
AC Q9VHK4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PYD protein.
GN PYD OR CG3763.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck C.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry N.C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```

```

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003680; AAF54300.1; -.
DR HSSP; Q12923; 3PDZ..
DR FlyBase; FBgn00031177; pyd.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR000906; ZU5.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00791; ZU5; 1.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00218; ZU5; 1.
DR PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 6.
DR SEQUENCE 1445 AA; 156723 MW; A5675FB66676A40E CRC64;

Query Match      25.3%; Score 97.5; DB 5; Length 1445;
Best Local Similarity 31.9%; Pred. No. 0.00045;
Matches 22; Conservative 13; Mismatches 31; Indels 3; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAKELQVHCNEVLHVTDTMQG-CGCWHAHRVNSYTMKDTAAHGT 59
||||: : : :|||: :| |||: |||: :|
DB 498 GDSFHKTHFCHDNPCKGEMAFKAGDVRVIDTLHNGVVGSMQVLKIGRGHQE--MQRGV 555

QY 60 IPNYSRAQ 68
||| |||::
DB 556 IPNKSRAEE 564

RESULT 12
Q96KB4 PRELIMINARY; PRT; 548 AA.
AC Q96KB4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA FLJ14386 fis, clone HEMBA1002417, moderately similar to TIGHT
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO, AND MAINLY HEAD;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027292; BAB55020.1; -.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR01597; ZONOCCLUDNS.
DR PRINTS; PR01600; ZONOCCLUDNS3.
DR PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 1.

```

```
DR PROSITE; PS50002; SH3; 1.
SQ SEQUENCE 548 AA; 61005 MW; ALA18ECFDE2627AC CRC64;

Query Match      25.0%; Score 96.5; DB 4; Length 548;
Best Local Similarity 38.4%; Pred. No. 0.00019;
Matches 28; Conservative 7; Mismatches 33; Indels 5; Gaps 2;

Qy 1 GDSFYRIVNLAMEGRAKGELOVHCNEVLHVDTMTFQCGCWH---HRVNSYTMKD--TA 55
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 104 GDSFYRTHFEELQPPSGLGTRGDVPHVLTDLHPGQSHARGGHHVLAVRMGRDLREQ 163
   ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy 56 AHGTPNYSRAQQ 68
   | ||| ||| : |
Db 164 ERGIIPNQSRAEQ 176
   | ||| ||| : |

RESULT 13
Q9VKG8
ID Q9VKG8 PRELIMINARY; PRT; 1916 AA.
AC
DT 01-MAY-2000 (TrEMBLrel. 13; Created)
DT 01-MAY-2000 (TrEMBLrel. 13; Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21; Last annotation update)
DE CG6509 protein (LD32687p).
GN CG6509.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003632; AAF53102.1; -.
DR EMBL; AY069586; AAL39731.1; -.
DR HSSP; P31016; 1BE9.
DR FlyBase; FBgn0032363; CG6509.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 4.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 4.
SQ SEQUENCE 1916 AA; 209863 MW; 32B2A61ABA6848F8 CRC64;

Query Match      24.6%; Score 95; DB 5; Length 1916;
Best Local Similarity 34.2%; Pred. No. 0.0014;
Matches 26; Conservative 14; Mismatches 24; Indels 12; Gaps 4;

Qy 1 GDSFYRIVNLAMEGRA-KGELOVHCNEVLHVDTMTFQGC-GCWHHRVNSYTMKDAAH- 57
   ||||| | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1590 GDSFYIRVGFDRTGELNEDDLRFVKNDEVLYVDNVFNCTFGLWRANKL-----DAMGHR 1643
   ||||| | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy 58 ---GTIPNYSRAQQOL 70
   | || : : : : |
Db 1644 KCGIIPSQMKVEBEL 1659
   | || : : : : |

RESULT 14
Q9UE73
ID Q9UE73 PRELIMINARY; PRT; 674 AA.
AC
DT 01-MAY-2000 (TrEMBLrel. 13; Created)
DT 01-MAY-2000 (TrEMBLrel. 13; Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21; Last annotation update)
DE Discs large protein P-dlg.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98409314; PubMed=9738934;
RA Nakamura H., Sudo T., Tsuiki H., Miyake H., Morisaki T., Sasaki J.,
RA Masuko N., Kochi M., Ushio Y., Saya H.;
RT "Identification of a novel human homolog of the Drosophila dlg, P-dlg,
RT specifically expressed in the gland tissues and interacting with
RT p55.";
RL FEBS Lett. 433:63-67(1998).
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; U61843; AAC61295.1; -.
DR HSSP; Q12959; 1PDR.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR000408; Reg_chrom_condens.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 2.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 2.
```

```
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 674 AA; 74892 MW; C4C006B0A4F7E8E8 CRC64;

Query Match      20.3%; Score 78.5; DB 4; Length 674;
Best Local Similarity 34.9%; Pred. No. 0.069;
Matches 22; Conservative 10; Mismatches 28; Indels 3; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAGKELQVHCNEVLHVDTDTMFQGC-GCWHHRVNSYTMKDTAAHGT 59
   ||||| : || : || : || : || : || : || : || : || : || : || : || : ||
Db 348 GDSFYIRALYDLADVEQELSFKKDDILYVDDTLPGTFGSMMWQLDENAKQ--IQRGQ 405
   ||| : || : || : || : || : || : || : || : || : || : || : || : ||

QY 60 IPN 62
   ||:
Db 406 IPS 408
   ||:

RESULT 15
Q9Y4E3
ID Q9Y4E3 PRELIMINARY; PRT; 1281 AA.
AC Q9Y4E3;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE KIAA0583 protein (Fragment).
GN KIAA0583.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Chara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB011155; BAA25509.1; -.
DR HSSP; Q12959; 1PDR.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR000408; Reg_chromatens.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 4.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00228; PDZ; 4.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 2.
DR PROSITE; PS50106; PDZ; 4.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
DR PROSITE; PS50002; SH3; 2.
KW SH3 domain.
FT NON_TER 1
SQ SEQUENCE 1281 AA; 140286 MW; 4CC150D03F3ED8B9 CRC64;

Query Match      20.3%; Score 78.5; DB 4; Length 1281;
Best Local Similarity 34.9%; Pred. No. 0.15;
Matches 22; Conservative 10; Mismatches 28; Indels 3; Gaps 2;

QY 1 GDSFYIRVNLAMEGRAGKELQVHCNEVLHVDTDTMFQGC-GCWHHRVNSYTMKDTAAHGT 59
   ||||| : || : || : || : || : || : || : || : || : || : || : || : ||
Db 955 GDSFYIRALYDLADVEQELSFKKDDILYVDDTLPGTFGSMMWQLDENAKQ--IQRGQ 1012
   ||| : || : || : || : || : || : || : || : || : || : || : || : ||

QY 60 IPN 62
   ||:
```

Db 1013 IPS 1015

Search completed: January 22, 2003, 08:56:45
Job time : 13.914 secs

THIS PAGE BLANK (user)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 34.8857 Seconds
(without alignments)
683.714 Million cell updates/sec

Title: US-09-767-215-2_COPY_826_1004
Perfect score: 930
Sequence: 1 RPRVLLVPRAGKILSEKL.....VROAIADEQKKVVWTEQSPR 179

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				A_Geneseq_101002:*	
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*	2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*	3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*	5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*	6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*	8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*	9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*	11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*	12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*	14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*	15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*	17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*	18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*	20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*	21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*	23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	930	100.0	1004	22 AAE07164	Human caspase recr
2	884	95.1	1139	22 AAE07165	Human predicted ca
3	242.5	26.1	1147	22 AAU01207	Human caspase recr
4	175.5	18.9	350	22 AAM38934	Human polypeptide
5	175.5	18.9	746	22 AAM39382	Human polypeptide
6	175.5	18.9	1032	22 AAU73247	Human plakoglobin
7	169.5	18.2	1032	22 AAU01206	Human caspase recr
8	138.5	13.8	674	19 AAU72748	Human P-dlg protei
9	124.5	13.4	198	16 AAR93137	Mouse guanylate ki
10	121.5	13.1	197	8 AAP70243	AA sequence (iii)

11	121.5	13.1	198	20 AAY07443	Mouse guanylate ki
12	121.5	13.1	198	20 AAP97135	Mouse guanylate ki
13	119.5	12.8	197	9 AAP80478	Sequence of a poly
14	119.5	12.8	197	16 AAR93136	Human guanylate ki
15	119.5	12.8	197	23 AAG15490	Human guanylate ki
16	118.5	12.7	197	20 AAY07442	Human guanylate ki
17	118.5	12.7	197	20 AAP97134	Human guanylate ki
18	117.5	12.6	236	9 AAP80626	AA sequence (VI) o
19	113.5	12.2	191	8 AAP70247	AA sequence (VII)
20	113.5	12.2	191	9 AAP80627	AA sequence (VII)
21	113.5	12.2	194	8 AAP70244	AA sequence (IV) o
22	113.5	12.2	194	9 AAP80624	AA sequence (IV) o
23	113.5	12.2	202	8 AAP70245	AA sequence (V) of
24	113.5	12.2	202	9 AAP80625	AA sequence (V) of
25	113.5	12.2	224	8 AAP70248	AA sequence (VIII)
26	113.5	12.2	224	9 AAP80628	AA sequence (VIII)
27	113.5	12.2	236	8 AAP70246	AA sequence (VI) o
28	110.5	11.9	1445	22 AB864619	Drosophila melanog
29	106	11.4	1916	22 AB862423	Drosophila melanog
30	106	11.4	1916	22 AB866089	Drosophila melanog
31	103.5	11.1	1736	23 AAU84308	Human endometrial
32	100.5	10.8	1745	23 AB857253	Mouse ischaemic co
33	94.5	10.2	1037	22 ABG22366	Novel human diagno
34	92.5	9.9	233	22 AB865915	Drosophila melanog
35	91	9.8	817	19 AAU48101	Human discs large
36	91	9.8	817	20 AAY30137	Amino acid sequenc
37	91	9.8	849	19 AAU48102	Human discs large
38	86.5	9.3	798	20 AAY01688	Protein encoded by
39	86.5	9.3	875	20 AAY01686	Protein encoded by
40	86.5	9.3	1290	20 AAY06809	Human Per gene pro
41	86.5	9.3	1290	21 AAY32216	Human PER protein.
42	86.5	9.3	1290	23 AB809289	Human period (Dros
43	86.5	9.3	1291	20 AAY01687	Protein encoded by
44	85	9.1	428	22 ABG10023	Novel human diagno
45	84	9.0	368	22 AB870791	Drosophila melanog

ALIGNMENTS

RESULT 1				AAE07164	
ID	AAE07164	standard; Protein; 1004 AA.			
XX	AAE07164;				
XX	AAE07164;				
DT	06-NOV-2001	(first entry)			
XX	Human caspase recruitment domain-14 (CARD-14).				
XX	Human: caspase recruitment domain-14; CARD-14; chromosome 17;				
KW	nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy;				
KW	autoimmune disorder; systemic lupus erythematosus; neurological disorder;				
KW	Alzheimer's disease; Parkinson's disease; inflammatory disorder; anaemia;				
KW	haematological disorder; myelodysplastic syndrome; myocardial infection;				
KW	stroke; immune disorder; Crohn's disease; allergic rhinitis; infection;				
KW	cell signalling disorder; cytostatic; immunosuppressive; nootropic;				
KW	neuroprotective; antiviral; antibacterial.				
OS	Homo sapiens.				
PH	Key	Location/Qualifiers			
FT	Modified-site	6..9			
FT	/note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"				
FT	Domain	10..116			
FT	/label= CARD_domain				
FT	Modified-site	12..15			
FT	/note= "Casein kinase II phosphorylation site"				
FT	Modified-site	18..21			
FT	/note= "Casein kinase II phosphorylation site"				
FT	Modified-site	25..27			
FT	/note= "Protein kinase C phosphorylation site"				

FT	Modified-site	60..62	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	/note= "N-glycosylation site"
FT	Modified-site	91..93	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	602..605
FT	Modified-site	114..117	/note= "N-glycosylation site"	FT	Modified-site	634..637
FT	Modified-site	117..122	/note= "N-myristoylation site"	FT	Modified-site	653..655
FT	Modified-site	121..123	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	674..677
FT	Domain	126..420	/label= Coiled_Coil_domain	FT	Domain	676..745
FT	Modified-site	130..135	/note= "N-myristoylation site"	FT	Modified-site	714..719
FT	Modified-site	134..137	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	725..727
FT	Modified-site	161..166	/note= "N-myristoylation site"	FT	Modified-site	725..728
FT	Modified-site	165..168	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	733..738
FT	Modified-site	220..227	/note= "Tyrosine kinase phosphorylation site"	FT	Modified-site	737..740
FT	Modified-site	221..224	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	759..761
FT	Domain	239..325	/label= k-Box_domain	FT	Modified-site	760..763
FT	Modified-site	240..243	/note= "Casein kinase II phosphorylation site"	FT	Peptide	785..793
FT	Modified-site	250..252	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	796..799
FT	Modified-site	253..256	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	809..805
FT	Modified-site	259..262	/note= "Casein kinase II phosphorylation site"	FT	Domain	826..1004
FT	Modified-site	280..283	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	/label= Guanylate_kinase_domain
FT	Modified-site	290..293	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	842..844
FT	Modified-site	297..300	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	860..863
FT	Modified-site	307..309	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	868..870
FT	Modified-site	307..310	/note= "Casein kinase II phosphorylation site"	FT	Region	870..872
FT	Modified-site	359..365	/note= "Tyrosine kinase phosphorylation site"	FT	Modified-site	893..896
FT	Modified-site	366..368	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	926..929
FT	Modified-site	366..369	/note= "Casein kinase II phosphorylation site"	FT	Peptide	941..949
FT	Modified-site	378..381	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	944..947
FT	Modified-site	384..386	/note= "Protein kinase C phosphorylation site"	FT	Modified-site	976..979
FT	Region	385..406	/note= "Leucine zipper pattern"	FT	Modified-site	980..985
FT	Modified-site	449..452	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	1002..1004
FT	Modified-site	463..466	/note= "Casein kinase II phosphorylation site"	FT	Modified-site	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	463..465	/note= "Protein kinase C phosphorylation site"	FT	WO200159065-A2.	
FT	Modified-site	470..472	/note= "Protein kinase C phosphorylation site"	FT	16-AUG-2001.	
FT	Modified-site	501..504	/note= "Casein kinase II phosphorylation site"	FT	22-JAN-2001; 2001WO-US02087.	
FT	Modified-site	511..516	/note= "N-myristoylation site"	FT	09-FEB-2000; 2000US-0181159.	
FT	Domain	568..660	/label= PDZ_domain	FT	(MILL-) MILLENNIUM PHARM INC.	
FT	Modified-site	587..592	/note= "N-myristoylation site"	FT	Bertin J;	
FT	Modified-site	589..592		FT	WPI; 2001-497073/54.	
				FT	N-PSDB; AAD13447.	
				FT	An isolated caspase recruitment domain polypeptide useful for	

PT regulating growth and cell death and useful for the treatment of cancer
XX
XX
PS
XX
XX
CC The present sequence is human caspase recruitment domain-14 (CARD-14).
CC The CARD-14 gene is located on chromosome 17. The CARD-14 is used for
CC the detection of modulators that modulates the ability of CARD-14 to
CC bind to Bcl-10 and stimulates phosphorylation of Bcl-10 or activation
CC of nuclear factor-kappa B (NF-kB). The CARD-14 is useful for regulating
CC growth and cell death and useful for the treatment of cancer. It is
CC also useful for the treatment of autoimmune disorders (e.g., systemic
CC lupus erythematosus), neurological disorders e.g., Alzheimer's and
CC Parkinson's disease, inflammatory disorders, haematological disorders
CC (e.g., anaemia, myelodysplastic syndromes), myocardial infarctions,
CC strokes, immune disorders (e.g., Crohn's disease, allergic rhinitis),
CC cell signalling disorders and certain viral and bacterial infections.
XX
SQ Sequence 1004 AA;
Query Match 100.0%; Score 930; DB 22; Length 1004;
Best Local Similarity 100.0%; Pred. No. 2.2e-91;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RRPVLLVPRVAGKILSEKILCLLQGFKKCLAEYLSQEEYEAWNSQSGDIIOEGVSGRCW 60
Db 826 RRPVLLVPRVAGKILSEKILCLLQGFKKCLAEYLSQEEYEAWNSQSGDIIOEGVSGRCW 885
QY 61 VTRHAVESLMEKNTALLDVLQDSVCTLHRMDIFPIVHVSVNEKMAKKLKGLORLGTS 120
Db 886 VTRHAVESLMEKNTALLDVLQDSVCTLHRMDIFPIVHVSVNEKMAKKLKGLORLGTS 945
QY 121 EEQLLEAARQEEGDLDRAPCLYSSSLAPDQWSLDGLLSCVROAIADQKVVYVTEQSPR 179
Db 946 EEQLLEAARQEEGDLDRAPCLYSSSLAPDQWSLDGLLSCVROAIADQKVVYVTEQSPR 1004
RESULT 2
AAE07165
ID AAE07165 standard; Protein; 1139 AA.
XX
AC AAE07165;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human predicted caspase recruitment domain-14 (CARD-14).
XX
KW Human; caspase recruitment domain-14; CARD-14; chromosome 17;
KW nuclear factor-kappa B; NF-kB; cell growth; cell death; cancer; therapy;
KW autoimmune disorder; systemic lupus erythematosus; neurological disorder;
KW Alzheimer's disease; Parkinson's disease; inflammatory disorder; anaemia;
KW haematological disorder; myelodysplastic syndrome; myocardial infarction;
KW stroke; immune disorder; Crohn's disease; allergic rhinitis; infection;
KW cell signalling disorder; cytostatic; immunosuppressive; neurotropic;
KW neuroprotective; antiviral; antibacterial.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 700
FT /note= "Encoded by TGG"
XX
XX WO200159065-A2.
XX
XX 16-AUG-2001.
XX
XX 22-JAN-2001; 2001WO-US02087.
XX
XX 09-FEB-2000; 2000US-0181159.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Bertin J;

XX WPI; 2001-497073/54.
DR N-PSDB; AAD13448.
XX
XX
PT An isolated caspase recruitment domain polypeptide useful for
PT regulating growth and cell death and useful for the treatment of cancer
PT
XX
PS Disclosure; Fig 2A-2C; 109pp; English.
XX
CC The present sequence is predicted human caspase recruitment domain-14
CC (CARD-14). The CARD-14 gene is located on chromosome 17. The CARD-14 is
CC used for the detection of modulators that modulates the ability of
CC CARD-14 to bind to Bcl-10 and stimulates phosphorylation of Bcl-10 or
CC activation of nuclear factor-kappa B (NF-kB). The CARD-14 is useful for
CC regulating growth and cell death and useful for the treatment of cancer.
CC It is also useful for the treatment of autoimmune disorders (e.g.,
CC systemic lupus erythematosus), neurological disorders e.g., Alzheimer's
CC and Parkinson's disease, inflammatory disorders, haematological disorders
CC (e.g., anaemia, myelodysplastic syndromes), myocardial infarctions,
CC strokes, immune disorders (e.g., Crohn's disease, allergic rhinitis),
CC cell signalling disorders and certain viral and bacterial infections.
XX
SQ Sequence 1139 AA;
Query Match 95.1%; Score 884; DB 22; Length 1139;
Best Local Similarity 100.0%; Pred. No. 2.4e-86;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RRPVLLVPRVAGKILSEKILCLLQGFKKCLAEYLSQEEYEAWNSQSGDIIOEGVSGRCW 60
Db 876 RRPVLLVPRVAGKILSEKILCLLQGFKKCLAEYLSQEEYEAWNSQSGDIIOEGVSGRCW 935
QY 61 VTRHAVESLMEKNTALLDVLQDSVCTLHRMDIFPIVHVSVNEKMAKKLKGLORLGTS 120
Db 936 VTRHAVESLMEKNTALLDVLQDSVCTLHRMDIFPIVHVSVNEKMAKKLKGLORLGTS 995
QY 121 EEQLLEAARQEEGDLDRAPCLYSSSLAPDQWSLDGLLSCVROAIADQKVV 171
Db 996 EEQLLEAARQEEGDLDRAPCLYSSSLAPDQWSLDGLLSCVROAIADQKVV 1046
RESULT 3
AAU01207
ID AAU01207 standard; Protein; 1147 AA.
XX
AC AAU01207;
XX
DT 12-SEP-2001 (first entry)
XX
DE Human caspase recruitment domain, CARD-11 polypeptide.
XX
KW Human; caspase recruitment domain; CARD-11; Bcl-10; NF-kappaB;
KW apoptosis; hyperproliferative disorder; autoimmune; neurological;
KW inflammatory disorder; viral infection; stress-related response.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Domain 6..112
FT /note= "CARD domain"
FT Modified-site 7..9
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 7..10
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 100..102
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 100..103
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 105..107
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 106..109
FT /note= "cAMP- and cGMP-dependent protein kinase

FT phosphorylation site"
 FT 130..431
 FT /note= "Coiled coil domain"
 FT 162..165
 FT /note= "casein kinase II phosphorylation site"
 FT 168..171
 FT /note= "Casein kinase II phosphorylation site"
 FT 175..183
 FT /note= "Tyrosine kinase phosphorylation site"
 FT 182..185
 FT /note= "Casein kinase II phosphorylation site"
 FT 189..195
 FT /note= "Tyrosine kinase phosphorylation site"
 FT 241..244
 FT /note= "N-glycosylation site"
 FT 243..245
 FT /note= "Protein kinase C phosphorylation site"
 FT 282..285
 FT /note= "Amidation site"
 FT 286..289
 FT /note= "Casein kinase II phosphorylation site"
 FT 290..292
 FT /note= "Protein kinase C phosphorylation site"
 FT 378..381
 FT /note= "Casein kinase II phosphorylation site"
 FT 429..432
 FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
 FT 459..461
 FT /note= "Protein kinase C phosphorylation site"
 FT 471..474
 FT /note= "Casein kinase II phosphorylation site"
 FT 472..475
 FT /note= "N-glycosylation site"
 FT 476..479
 FT /note= "Casein kinase II phosphorylation site"
 FT 508..510
 FT /note= "Protein kinase C phosphorylation site"
 FT 510..513
 FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
 FT 538..560
 FT /note= "Protein kinase C phosphorylation site"
 FT 578..581
 FT /note= "Casein kinase II phosphorylation site"
 FT 584..587
 FT /note= "N-glycosylation site"
 FT 587..592
 FT /note= "N-myristoylation site"
 FT 634..637
 FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
 FT 635..748
 FT /note= "PDZ domain"
 FT 635..1147
 FT /note= "MAGUK domain"
 FT 638..641
 FT /note= "Glycosaminoglycan attachment site"
 FT 678..683
 FT /note= "N-myristoylation site"
 FT 687..689
 FT /note= "Protein kinase C phosphorylation site"
 FT 692..695
 FT /note= "Casein kinase II phosphorylation site"
 FT 698..703
 FT /note= "N-myristoylation site"
 FT 710..715
 FT /note= "N-myristoylation site"
 FT 725..728
 FT /note= "Casein kinase II phosphorylation site"
 FT 761..766
 FT /note= "N-myristoylation site"
 FT 764..767

FT Domain
 FT /note= "Casein kinase II phosphorylation site"
 FT 766..834
 FT /note= "SH3 domain"
 FT 776..779
 FT /note= "N-glycosylation site"
 FT 779..782
 FT /note= "Casein kinase II phosphorylation site"
 FT 787..789
 FT /note= "Protein kinase C phosphorylation site"
 FT 816..819
 FT /note= "Casein kinase II phosphorylation site"
 FT 823..828
 FT /note= "N-myristoylation site"
 FT 847..850
 FT /note= "Casein kinase II phosphorylation site"
 FT 853..858
 FT /note= "N-myristoylation site"
 FT 857..859
 FT /note= "Protein kinase C phosphorylation site"
 FT 872..875
 FT /note= "Casein kinase II phosphorylation site"
 FT 882..1147
 FT /note= "Guanylate kinase (GUK) domain"
 FT 897..900
 FT /note= "Casein kinase II phosphorylation site"
 FT 917..922
 FT /note= "N-myristoylation site"
 FT 926..929
 FT /note= "Casein kinase II phosphorylation site"
 FT 935..937
 FT /note= "Protein kinase C phosphorylation site"
 FT 1003..1006
 FT /note= "Casein kinase II phosphorylation site"
 FT 1010..1018
 FT /note= "Tyrosine kinase phosphorylation site"
 FT 1050..1055
 FT /note= "N-myristoylation site"
 FT 1088..1091
 FT /note= "Casein kinase II phosphorylation site"
 FT 1120..1123
 FT /note= "Casein kinase II phosphorylation site"
 FT WO200140468-A2.
 XX
 PN 07-JUN-2001.
 XX PD
 XX
 PF 01-DEC-2000; 2000WO-US32716.
 XX
 PR 03-DEC-1999; 990US-0168780.
 PR 18-FEB-2000; 2000US-0507533.
 PR 25-FEB-2000; 2000US-0513904.
 PR 10-OCT-2000; 2000US-0685791.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 PI Bertin J;
 XX
 DR WPI: 2001-367809/38.
 DR N-PSDB; AAS05389.
 XX
 PT Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10, CARD-11, useful as targets for therapy, as immunogens, and in screening and detection assays -
 PT
 XX Claim 9; Fig 14A-14C; 145pp; English.
 PS
 XX The present sequence represents novel human caspase recruitment domain, CARD-11. The polynucleotide encoding this sequence was isolated from a human T-cell cDNA library. Also described are novel human sequences for CARD-9 and CARD-10 (AAU01205, AAU01206) and rat CARD-9 (AAU01204). CARD-9, CARD-10 and CARD-11 interact with Bcl-10 which is thought to activate nuclear factor (NF)-kappaB and apoptosis. The sequences of the invention can be used for

CC treating a disorder associated with abnormal levels of apoptosis by
CC modulating the expression or activity of CARD-9, CARD-10, or CARD-11.
CC They can be used for the treatment of hyperproliferative disorders
CC (e.g. cancer), autoimmune disorders (e.g. systemic lupus erythematosus),
CC neurological disorders (e.g. Alzheimer's disease), inflammatory disorders
CC (e.g. Crohn's disease), and viral infection (e.g. HIV). The CARD
CC polypeptide, polynucleotide and an antibody which selectively binds to
CC CARD can be used in screening and detection assays (e.g. chromosomal,
CC mapping, tissue typing), predictive medicine (prognostic assays,
CC monitoring clinical trials, and therapy (treatment and prophylaxis). The
CC CARD polypeptide may be used to screen for drugs that bind to and/or
CC modulate it. CARD sequences are potential targets for regulating
CC inflammation, cancer, NF-kappaB signalling, stress-related response and
CC apoptosis in human disease. A host cell containing a polynucleotide
CC encoding CARD can be used to create transgenic animals.
SQ Sequence 1147 AA;
Query Match 26.1%; Score 242.5; DB 22; Length 1147;
Best Local Similarity 31.5%; Pred. No. 5.8e-17;
Matches 57; Conservative 41; Mismatches 76; Indels 7; Gaps 4;
QY 1 RRPVLLVPRAVGKILSEKLCLOG---FKKCLAEVLSOEYEAWNSQSGDII--OGEVVS 55
DB 966 RRPVLLFTFTVIAKTLVORLLSGGAMETICKSDIVTDEFLRQKTTIYSREKNPN 1025
QY 56 GGRCWTRHAVESLMKNTALLDVLQDSVCTLHRMDIPPIVHVSNEKMAKILKKGLO 115
DB 1026 AFEC-IAPANIEAVAANKHCLLEAGIGCTRDLIKSNIYPIVLFIRVCEKNIKRFRKL 1084
QY 116 RLGTSSEQLLEAARQEGDLDRAPCLYSSLPDGSDDLGLLSCVRAIADQKVVWTE 175
DB 1085 RPET-EEEFVLCRLKEKEALPCLYATVPDMGVSVELLRVVKDTIGEQRKTIWVD 1143
QY 176 Q 176
DB 1144 E 1144
RESULT 4
AAM38934
ID AAM38934 standard; Protein; 350 AA.
AC AAM38934;
XX
XX
XX 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 2079.
XX Human; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
XX WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX

PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Mehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR N-PSDB; AAI58090.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Example 3; SEQ ID NO 2079; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX Sequence 350 AA;
SQ Query Match 18.9%; Score 175.5; DB 22; Length 350;
Best Local Similarity 32.8%; Pred. No. 2.1e-10;
Matches 58; Conservative 24; Mismatches 88; Indels 7; Gaps 3;
QY 3 RPVLLVPRAVGKILSEKLCLOG---FKKCLAEVLSOEYEAWNSQSGDIIOGEVVS 58
DB 164 RPVLLPECLAPRLIRNLLDLPSSRLDFQVCPAESLSGSELCPSSAPGAPKAQPTPGLG 223
QY 59 CWTTHAVESLMKNTALLDVLQDSVCTLHRMDIPPIVHVSNEKMAKILKKGLO 118
DB 224 SRI-RAIQESVGKK--HCLLELGARGVRELQVNEIYPIVHVEVTEKNVREVRLLGRFG 280
QY 119 TSEQLLEAARQEGDLDRAPCLYSSLPDGSDDLGLLSCVRAIADQKVVWTE 175
DB 281 WRDSELLROCRSEQVWLGLPCSWQVPAHENGHAEELAKVVRGRLIQEQARLVWVE 337
RESULT 5
AAM93822
ID AAM93822 standard; Protein; 746 AA.
XX
XX AAM93822;
XX 06-NOV-2001 (first entry)
XX Human polypeptide, SEQ ID NO: 3879.
XX Human; full length cDNA; cDNA synthesis; oligo-capping.
XX Homo sapiens.
XX EP1130094-A2.
XX 05-SEP-2001.
XX 07-JUL-2000; 2000EP-0114089.
XX 08-JUL-1999; 93JP-0194486.
XX 11-JAN-2000; 2000JP-0118774.
XX 02-MAY-2000; 2000JP-0183765.
XX

FT Modified-site 78..80 /note= "protein kinase C phosphorylation site"
 FT 88..91 /note= "Amidation site"
 FT Modified-site 112..115
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 113..118
 FT /note= "N-myristoylation site"
 FT Domain 147..457
 FT /note= "Coiled coil domain"
 FT Modified-site 201..207
 FT /note= "Tyrosine kinase phosphorylation site"
 FT Region 230..251
 FT /note= "Leucine zipper homology region"
 FT Modified-site 242..245
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 293..295
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 293..296
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 309..314
 FT /note= "N-myristoylation site"
 FT Modified-site 313..315
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 331..334
 FT /note= "Casein kinase II phosphorylation site"
 FT Domain 366..398
 FT /note= "Tropomyosin domain"
 FT Modified-site 412..415
 FT /note= "Casein kinase II phosphorylation site"
 FT Region 426..447
 FT /note= "Leucine zipper homology region"
 FT Modified-site 438..441
 FT /note= "Casein kinase II phosphorylation site"
 FT Domain 457..1032
 FT /note= "MAGUK domain"
 FT Modified-site 472..475
 FT /note= "N-glycosylation site"
 FT Modified-site 478..481
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 487..492
 FT /note= "N-myristoylation site"
 FT Modified-site 510..513
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 512..514
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 549..552
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 558..560
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 565..570
 FT /note= "N-myristoylation site"
 FT Modified-site 570..573
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 595..598
 FT /note= "N-glycosylation site"
 FT Modified-site 603..605
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 638..641
 FT /note= "Glycosaminoglycan attachment site"
 FT Modified-site 642..644
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 656..661
 FT /note= "N-myristoylation site"
 FT Modified-site 681..684
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 690..693
 FT /note= "Casein kinase II phosphorylation site"
 FT Domain 704..772
 FT /note= "SH3 domain"
 FT Modified-site 712..715
 FT /note= "N-glycosylation site"
 FT Modified-site 714..717

FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 733..739
 FT /note= "Tyrosine kinase phosphorylation site"
 FT Modified-site 748..751
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 754..756
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 754..757
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 761..766
 FT /note= "N-myristoylation site"
 FT Modified-site 782..784
 FT /note= "Protein kinase C phosphorylation site"
 FT Modified-site 809..814
 FT /note= "N-myristoylation site"
 FT Domain 830..1032
 FT /note= "Guanylate kinase (GUK) domain"
 FT Modified-site 830..832
 FT /note= "Protein kinase C phosphorylation site"
 FT Modified-site 868..870
 FT /note= "Protein kinase C phosphorylation site"
 FT Modified-site 869..872
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 882..885
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 893..898
 FT /note= "N-myristoylation site"
 FT Modified-site 915..918
 FT /note= "Amidation site"
 FT Modified-site 947..949
 FT /note= "Protein kinase C phosphorylation site"
 FT Modified-site 981..986
 FT /note= "N-myristoylation site"
 FT Modified-site 1021..1026
 FT /note= "N-myristoylation site"
 FT Modified-site 1022..1024
 FT /note= "Protein kinase C phosphorylation site"
 FT Modified-site 1028..1031
 FT /note= "Casein kinase II phosphorylation site"
 FT WO200140468-A2.
 PN 07-JUN-2001.
 PD 01-DEC-2000; 2000WO-US32716.
 PF 03-DEC-1999; 99US-0168780.
 PR 18-FEB-2000; 2000US-0507533.
 PR 25-FEB-2000; 2000US-0513904.
 PR 10-OCT-2000; 2000US-0685791.
 XX (MILL-) MILLENNIUM PHARM INC.
 PA Bertin J;
 PI WPI; 2001-367809/38.
 DR N-PSDB; AAS05388.
 XX Novel caspase recruitment domain (CARD) proteins, CARD-9, CARD-10,
 PT CARD-11, useful as targets for therapy, as immunogens, and in screening
 PT and detection assays -
 XX Claim 9; Fig 10A-10C; 145pp; English.
 PS The present sequence represents novel human caspase recruitment
 XX domain, CARD-10. The polynucleotide encoding this sequence was isolated
 CC from a human skin cDNA library. Also described are novel human sequences
 CC for CARD-9 and CARD-11 (AAU01205, AAU01207) and rat CARD-9 (AAU01204).
 CC CARD-9, CARD-10 and CARD-11 interact with Bcl-10 which is thought to
 CC activate nuclear factor (NF)-kappaB and apoptosis. The sequences of the
 CC invention can be used for treating a disorder associated with abnormal
 CC levels of apoptosis by modulating the expression or activity of CARD-9,
 CC CARD-10, or CARD-11. They can be used for the treatment of

CC intracellular parasitic diseases or bacterial infection.

XX SQ Sequence 198 AA;
Query Match 13.1%; Score 121.5; DB 20; Length 198;
Best Local Similarity 26.9%; Pred. No. 6.8e-05;
Matches 53; Conservative 37; Mismatches 78; Indels 29; Gaps 9;
QY 2 PRPVLLV-PRAVGK-ILSEKL-----CLLQGFKKCLAE-----YLSQEEVEA 41
Db 4 PRPVLSGSGAGKSTLLKLFQEHSSIFGFSVSHTRNPRGEEGDKDYFVTREMMQR 63
QY 42 WSQRGDIIEGEVSGRCWVTRHAVESLMENKTHALLDVLQDSVCTLRHMDIFFIVHVS 101
Db 64 DIAAGDFIEHAEFSGNLYGTSKEAVRAVQAMNRCVLDVLDLQAVRSIKKTDLCP--IYIF 121
QY 102 VNEKMAKKLKGQLRGT--SEQL---LEAARQEGDLDRAPCLYSSLPAGWSLDGLL 157
Db 122 VQPPSLDVLEQLRLRNTETESLAKRLPAARTDM-ESSKEPGLFDLVIIN--DLDKRAY 178
QY 158 SCVROAIADQOKKVVMT 174
Db 179 ATLKOALSEEIKKAQGT 195

RESULT 12
AAW971135
ID AAW971135 standard; Protein; 198 AA.
XX AC AAW971135;
XX DT 22-JUN-1999 (first entry)
XX DE Mouse guanylate kinase protein.

XX KW HSV-1; thymidine kinase; mutation; DRH nucleoside binding site; enzyme;
KW pathogen; tumour; hyperkeratosis; psoriasis; prostate hypertrophy; PCR;
KW hyperthyroidism; endocrinopathy; autoimmune disease; allergy; restenosis;
KW viral disease; AIDS; hepatitis; parasite; bacterial infection.
XX OS Mus sp.
XX PN US5877010-A.
XX PD 02-MAR-1999.
XX PF 02-MAY-1995; 95US-0432871.
XX PR 02-MAY-1995; 95US-0432871.
XX PR 02-MAY-1994; 94US-0237592.
XX PA (UNIW) UNIV WASHINGTON.
XX PI Black ME, Loeb LA;
XX WPI; 1999-189650/16.
XX DR N-PSDB; AAX15377.
XX New Herpesviridae thymidine kinase mutant nucleic acids - used to
PT develop products for treating e.g. tumours, autoimmune diseases,
PT allergies, restenosis or viral, bacterial or parasitic diseases
XX

Disclosure; Fig 22; 72pp; English.

XX This sequence represents the mouse guanylate kinase (GK) protein. The
XX invention relates to the generation of novel HSV-1 thymidine kinase (TK)
CC or GK genes with a mutation upstream, within or downstream from a DRH
CC nucleoside binding site. The TK enzymes can be used for inhibiting
CC pathogenic agents, e.g. tumours, hyperkeratosis, psoriasis, prostate
CC hypertrophy, hyperthyroidism, endocrinopathies, autoimmune diseases,
CC allergies, restenosis, viral diseases such as AIDS, hepatitis,
CC intracellular parasitic diseases or bacterial infection.

SQ Sequence 198 AA;
Query Match 13.1%; Score 121.5; DB 20; Length 198;
Best Local Similarity 26.9%; Pred. No. 6.8e-05;
Matches 53; Conservative 37; Mismatches 78; Indels 29; Gaps 9;
QY 2 PRPVLLV-PRAVGK-ILSEKL-----CLLQGFKKCLAE-----YLSQEEVEA 41
Db 4 PRPVLSGSGAGKSTLLKLFQEHSSIFGFSVSHTRNPRGEEGDKDYFVTREMMQR 63
QY 42 WSQRGDIIEGEVSGRCWVTRHAVESLMENKTHALLDVLQDSVCTLRHMDIFFIVHVS 101
Db 64 DIAAGDFIEHAEFSGNLYGTSKEAVRAVQAMNRCVLDVLDLQAVRSIKKTDLCP--IYIF 121
QY 102 VNEKMAKKLKGQLRGT--SEQL---LEAARQEGDLDRAPCLYSSLPAGWSLDGLL 157
Db 122 VQPPSLDVLEQLRLRNTETESLAKRLPAARTDM-ESSKEPGLFDLVIIN--DLDKRAY 178
QY 158 SCVROAIADQOKKVVMT 174
Db 179 ATLKOALSEEIKKAQGT 195

RESULT 13
AAP80478
ID AAP80478 standard; protein; 197 AA.
XX AC AAP80478;
XX DT 15-OCT-1990 (first entry)

XX DE Sequence of a polypeptide having human haematopoietic cell growth
XX potentiating factor (HCGPF) activity.
XX KW Human haematopoietic cell growth potentiating factor; immune control;
KW haematopoietic control functions; immunodeficient diseases;
KW autoimmune diseases; infectious diseases.
XX OS Homo sapiens.

XX FH Location/Qualifiers
XX FT Region 19..197
XX FT /note="A polypeptide with this sequence and a gene
FT encoding it are claimed"
FT Region 31..197
FT /note="As above"

XX EP274560-A.

XX PD 20-JUL-1988.

XX PF 19-JUN-1987; 87EP-0108782.

XX PR 08-JAN-1987; 87JP-0002521.

XX PA (AJTN) AJINOMOTO KK.

XX PI Taniguchi T, Yamada G, Hamuro J, Taki S, Matsui H, Kashima N;

XX WPI; 1988-199153/29.

XX DR N-PSDB; AAN81032.

XX PT Recombinant human haematopoietic cell growth potentiating factor -
PT used for immuno-deficient, auto-immune or infectious diseases,
PT hepatitis, nephritis, cancer or bone marrow transplantation

XX PS Claim 13, claim 34; Fig 2 and Fig 5 and Fig 8; 58pp; English.

XX A polypeptide with this sequence and a gene encoding it are claimed. Also
CC claimed is a vector DNA capable of replication in prokaryotes and
CC eukaryotes, the transformed prokaryotes and eukaryotes and a polypeptide
CC having HCGPF activity. The HCGPF potentiates the growth of haematopoietic
CC cells. It exhibits immune control and haematopoietic control functions.

```
CC used to inhibit a pathogenic agent in a warm blooded animal e.g. viruses;
XX bacteria or parasites, or a tumour cell or autoreactive immune cell.
XX SQ Sequence      197 AA;

Query Match          12.8%; Score 119.5; DB 16; Length 197;
Best Local Similarity 25.5%; Pred. No. 0.00011;
Matches 49; Conservative 38; Mismatches 78; Indels 27; Gaps 7

QY    2 PRPVLLV-PRAVGKILSEKLCLLQ-----GPKKKLAEE-----YLSEQEYEAA 41
Db     ||||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY    4 PRPVLPSPSCAGASTLKRLQLQEHSGIFGSVSHTNRPGRCEENGKDYYFVTREVMOR 63
QY   42 WSQLGDIIQEVEVGSGRWTRHVAVESLMEXKNTHALLDVQLDSVCTLHRMDIFFPIVIHVS 101
Db     || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   64 DIAAGDFIEHAEFSGNLYGTSKVAOVAQMNRICVDLDVQGVRNKATDLRP--IYIS 121
QY   102 VNERMAKKLKGQLQRIGT-SSEQLLE--AARQEGDLDRAPCLYSLSAPDGWSDLDGLLS 158
Db     ||| :::: | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   122 VPQPSHLHVLEQLRQRNTETEESLVKRLLAAAQAADMESKEPGLFDVVIIIN--DSLDAQAYA 179
QY   159 CVROAIADQOKK 170
Db     :::|:::| | |

DB   180 ELKEALSEEIKK 191

RESULT 15
ID AAO15490 standard; Protein; 197 AA.
XX AC AAO15490;
XX DT (first entry)
XX DE Human guanylate kinase 1 (GUK1).
XX KW Human; DHAM-kinase inhibitor; guanylate kinase 1; PRK2;
KW deregulated in hyperactive macrophage kinase inhibitor; GUK1;
KW serin-threonin-kinase; PAK2; inflammatory condition;
KW chronic inflammatory airway disease; chronic bronchitis;
KW chronic obstructive pulmonary disease; COPD.
XX OS Homo sapiens.
XX WO200252036-A2.
XX PD 04-JUL-2002.
XX PF 15-DEC-2001; 2001WO-EPI4844.
XX PR N-PADB; AAL44148.
XX PA (BOEH ) BOEHRINGER INGELHEIM PHARMA KG.
XX PI Jung B, Mueller S, Kraut N;
XX WIPI; 2002-583570/62.
XX DR N-PSDB; AAL44148.
XX PT Determining activators or inhibitors of 'deregulated in hyperactive
PT macrophage' (DHAM)-kinase for treating chronic inflammatory airway
PT diseases, by measuring DHAM-kinase function after it is contacted with
PT a test substance -
XX PS Claim 8; Page 47-48; 48pp; English.
XX The invention comprises a method for determining whether a substance is
CC an activator or an inhibitor of a DHAM-kinase (deregulated in hyperactive
CC macrophage kinase). DHAM-kinaes used in the invention include guanylate
CC kinase 1 (GUK1), serin-threonin-kinase PAK2 and serin-threonin-kinase
CC PKK2. The method of the invention is useful for identifying substances
CC that influence inflammatory conditions of chronic inflammatory airway
CC diseases (e.g. chronic bronchitis or chronic obstructive pulmonary
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:52:30 ; Search time 12.2702 Seconds
(without alignments)
429.229 Million cell updates/sec

Title: US-09-767-215-2_COPY_826_1004

Perfect score: 930

Sequence: 1 RPRVLLVPRVAGKILSEKL.....VROAIADQKVVWTEQSPR 179

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	121.5	13.1	198	2	US-08-432-871C-51
2	121.5	13.1	198	4	US-09-270-956-51
3	118.5	12.7	197	2	US-08-432-871C-49
4	118.5	12.7	197	4	US-09-270-956-49
5	86.5	9.3	798	4	US-09-150-460B-8
6	86.5	9.3	875	4	US-09-150-460B-7
7	86.5	9.3	1290	4	US-09-150-460B-6
8	84	9.0	724	4	US-09-562-737-21
9	83	8.9	724	4	US-09-562-737-29
10	81	8.7	2482	1	US-08-328-254-6
11	80	8.6	724	4	US-09-562-737-24
12	80	8.6	3248	1	US-08-353-700-1
13	80	8.6	3248	5	PCT-US95-16216-1
14	76.5	8.2	460	1	US-08-289-709-1
15	76.5	8.2	460	1	US-08-602-656-1
16	76.5	8.2	1291	4	US-09-150-460B-10
17	76.5	8.2	1291	4	US-09-220-641-5
18	75	8.1	647	1	US-07-894-212A-8
19	75	8.1	649	1	US-07-894-212A-2
20	75	8.1	650	1	US-07-893-928A-1
21	75	8.1	674	4	US-08-961-083-200
22	73.5	7.9	1048	4	US-08-887-534A-85
23	73	7.8	1162	4	US-09-134-001C-4008
24	72.5	7.8	349	4	US-09-161-241-8
25	72.5	7.8	724	4	US-09-562-737-27
26	72.5	7.8	976	4	US-09-302-812-4
27	72.5	7.8	976	4	US-09-511-477-4

28	72.5	7.8	976	4	US-09-511-507-4	Sequence 4, Appli
29	72	7.7	724	4	US-09-562-737-23	Sequence 23, Appl
30	71.5	7.7	724	4	US-09-562-737-28	Sequence 28, Appl
31	71.5	7.7	1094	2	US-08-680-326-40	Sequence 40, Appl
32	71	7.6	490	1	US-08-361-611-2	Sequence 2, Appli
33	71	7.6	490	1	US-08-565-655-2	Sequence 2, Appli
34	71	7.6	490	2	US-08-946-967-2	Sequence 2, Appli
35	70.5	7.6	186	1	US-08-117-083-22	Sequence 22, Appl
36	70.5	7.6	1066	4	US-09-541-782-8	Sequence 8, Appli
37	70.5	7.6	1066	4	US-09-723-820-8	Sequence 8, Appli
38	70	7.5	351	4	US-09-134-001C-5492	Sequence 5492, Ap
39	70	7.5	386	4	US-09-085-199B-2	Sequence 2, Appli
40	70	7.5	724	4	US-09-562-737-30	Sequence 30, Appl
41	70	7.5	914	4	US-09-085-199B-4	Sequence 4, Appli
42	70	7.5	1079	3	US-09-058-489-22	Sequence 22, Appl
43	70	7.5	1240	3	US-09-058-489-23	Sequence 23, Appl
44	70	7.5	1347	3	US-09-058-489-24	Sequence 24, Appl
45	69.5	7.5	250	4	US-09-036-987A-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-08-432-871C-51
; Sequence 51, Application US/08432871C
; Patent No. 5877010
; GENERAL INFORMATION:
; APPLICANT: Loeb, Lawrence A.
; APPLICANT: Black, Margaret E.
; TITLE OF INVENTION: THYMIDINE KINASE MUTANTS
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432.871C
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 240052.409C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 198 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-432-871C-51

Query Match	13.1%	Score	121.5	DB	2	Length	198
Best Local Similarity	26.9%	Pred. No.	1e-05				
Matches	53	Conservative	37	Mismatches	78	Indels	29
Gaps	9						
Oy	2	PRPVLLV-PRVAGK-ILSEKL-----CLLQFKKCLAE-----	YLSQEEVEA	41			
Db	4	PRPVLLSPGAGKSTLKKLFQEHSSIFGFSVHTNRPGRGEEDGKYFVETREMMQR	63				
Oy	42	WSQRDITIQEGEVSGGRCWVTRHVESLMKNTKTHALLDVLQDSVCTLHRMDFPIVHVS	101				

```
Db 64 DIAAGDFIEHAESGNYLGTSGEAVRAVQAMNRCVLDVLDQAVRSIKKTDLCP--IYIF 121
Qy 102 VNEKMAKLLKGLQRLGT--SEQL---LEAARQEGDLDRAPCLYSSSLAPGWSLDGLL 157
Db 122 VOPPSLDVLEQPLRLRNTETESLAKRLPAARTDM--ESSKEPGLFDLVIIN--DDLDKAY 178
Qy 158 SCVROAIADQKVVWT 174
Db 179 ATLQALSEEIKKAQGT 195

RESULT 2
US-09-270-956-51
; Sequence 51, Application US/09270956
; Patent No. 6451571
; GENERAL INFORMATION:
; APPLICANT: Loeb, Lawrence A.
; APPLICANT: Black, Margaret E.
; TITLE OF INVENTION: THYMIDINE KINASE MUTANTS
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/270,956
; FILING DATE: 17-MAR-1999
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 240052.409C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 198 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-270-956-51

Query Match 13.1%; Score 121.5; DB 4; Length 198;
Best Local Similarity 26.9%; Pred. No. 1e-05;
Matches 53; Conservative 37; Mismatches 78; Indels 29; Gaps 9;

Qy 2 PRPVLLV-PRAVGK-ILSEKL-----CLQGFKKCLAE-----YLSQEEYEA 41
Db 4 PRPVLSGSGAGKSTLLKLLQEHSSIFGFSVSHTRNPRGEEGDKDYFVFTREMQR 63
Qy 42 WSQRGDIIEGEVSGRCWTRHAVESLMENKTHALLDVLQDSVCTLHRMDIFPIVHVS 101
Db 64 DIAAGDFIEHAESGNYLGTSGEAVRAVQAMNRCVLDVLDQAVRSIKKTDLCP--IYIF 121
Qy 102 VNEKMAKLLKGLQRLGT--SEQL---LEAARQEGDLDRAPCLYSSSLAPGWSLDGLL 157
Db 122 VOPPSLDVLEQPLRLRNTETESLAKRLPAARTDM--ESSKEPGLFDLVIIN--DDLDKAY 178
Qy 158 SCVROAIADQKVVWT 174
Db 179 ATLQALSEEIKKAQGT 195
```

```
RESULT 3
US-08-432-871C-49
; Sequence 49, Application US/08432871C
; Patent No. 5877010
; GENERAL INFORMATION:
; APPLICANT: Loeb, Lawrence A.
; APPLICANT: Black, Margaret E.
; TITLE OF INVENTION: THYMIDINE KINASE MUTANTS
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,871C
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 240052.409C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 197 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-432-871C-49

Query Match 12.7%; Score 118.5; DB 2; Length 197;
Best Local Similarity 25.5%; Pred. No. 2.2e-05;
Matches 49; Conservative 38; Mismatches 78; Indels 27; Gaps 7;

Qy 2 PRPVLLV-PRAVGKILSEKL-----LQGFKKCLAE-----YLSQEEYEA 41
Db 4 PRPVLSGSGAGKSTLLKLLQAHSGIFGFSVSHTRNPRGEEGDKDYFVFTREMQR 63
Qy 42 WSQRGDIIEGEVSGRCWTRHAVESLMENKTHALLDVLQDSVCTLHRMDIFPIVHVS 101
Db 64 DIAAGDFIEHAESGNYLGTSGKVAQVQAMNRCVLDVLDQGVNRNIKATDLRP--IYIS 121
Qy 102 VNEKMAKLLKGLQRLGT--SEQLLE---AARQEGDLDRAPCLYSSSLAPGWSLDGLL 158
Db 122 VOPPSLHVLEQRLRNTETESLVRKLAQAQADMESSKEPGLFDVVIIN--DSLQQA 179
Qy 159 CVROAIADQK 170
Db 180 ELKEALSEEIKK 191

RESULT 4
US-09-270-956-49
; Sequence 49, Application US/09270956
; Patent No. 6451571
; GENERAL INFORMATION:
; APPLICANT: Loeb, Lawrence A.
; APPLICANT: Black, Margaret E.
; TITLE OF INVENTION: THYMIDINE KINASE MUTANTS
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
```


RESULT 9

US-09-562-737-29

Sequence 29, Application US/09562737

Patent No. 6428967

GENERAL INFORMATION:

APPLICANT: Herz, Joachim

APPLICANT: Gotthardt, Michael

TITLE OF INVENTION: LDL Receptor Signaling Pathways

FILE REFERENCE: UTSW0708

CURRENT APPLICATION NUMBER: US/09/562,737

CURRENT FILING DATE: 2000-05-01

NUMBER OF SEQ ID NOS: 132

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 29

LENGTH: 724

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic

OTHER INFORMATION: Sequence

US-09-562-737-29

Query Match 8.9%; Score 83; DB 4; Length 724;

Best Local Similarity 18.7%; Pred. No. 1.2;

Matches 29; Conservative 32; Mismatches 68; Indels 26; Gaps

QY 35 SOEEYEAWSGDIIQGEYSGGRCVWTRHAVESLMKENTHALLDQLDSVCTLHRMDIF 94

Db 585 SREKKEDIQAHKFLQAGYNSHLMGTSVOSVRENABQGGKHCILQVSANAVRRLRAHLH 644

QY 95 PIVTHV-----SYNEKMAKL-KKGLQRIGTSEOLLEAAROEGLDRAPCLYSS 144

Db 645 PIAISIRPSLENVTEINKRITEEVARKFDRATWLQEFTE-----CFYAI 691

QY 145 LAPGWSDLGLSCVQRQATDEQKKVWTEQSPR 179

Db 692 VEGDSF---EAIYHKVRVIDLGGPYIWVEARER 723

RESULT 10

US-08-328-254-6

Sequence 6, Application US/08328254

Patent No. 5710022

GENERAL INFORMATION:

APPLICANT: Zhu, Xueliang

APPLICANT: Lee, Wen-Hwa

TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/328,254

FILING DATE: 24-OCT-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/141,239

FILING DATE: 22-OCT-1993

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-CJ 1191


```
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 460 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-602-636-1

Query Match      8.2%; Score 76.5; DB 1; Length 460;
Best Local Similarity 24.8%; Pred. No. 3.3;
Matches 32; Conservative 21; Mismatches 43; Indels 33; Gaps 4;

QY 27 KKCLAEYLSQEEYEAWSQSGDIQEGEVSGGRCWVTRHAVESLMKKNTHALLDVOLD SVC 86
Db 195 KKLAEGNTEPIYHALTRPEV--EGEATGRACQLTELAGSQL----- 235

QY 87 TLHRMDIFFPVIHVS---VNEKMAKKLKKGLQRLGTSEQ--LLEAARQEEGDLDRAPCL 141
Db 236 -----YVVHVTCAQAVEKIAQARNKGLDVWGGETCPQYLVLVDQSYLEKPDFEGAKYV 286

QY 142 YSSLAPD GW 150
Db 287 WSPPLREKW 295
```

Search completed: January 22, 2003, 08:57:42
Job time : 15.2702 secs

THIS PAGE BLANK (uspto)


```

; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-142001
; CURRENT APPLICATION NUMBER: US/09/767,215
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/181,159
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1138
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-767-215-5

Query Match          95.1%; Score 884; DB 10; Length 1138;
Best Local Similarity 100.0%; Pred. No. 4.2e-82;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRPVLLVPRVKGKILSEKLCILQGFKKCLAEYLSOEYEAWSQRGDIIOEGVSGRCW 60
Db 875 RRPVLLVPRVKGKILSEKLCILQGFKKCLAEYLSOEYEAWSQRGDIIOEGVSGRCW 934

QY 61 VTRHAVESLMKNTALLDVQLDSVCTLHRMDIFFPIVHVSNEKMAKLLKGLQRLGTS 120
Db 935 VTRHAVESLMKNTALLDVQLDSVCTLHRMDIFFPIVHVSNEKMAKLLKGLQRLGTS 994

QY 121 EQLLEAARQEGDLRAPCLYSSLPDGSWDLGLLSCVROAIADEQKV 171
Db 995 EQLLEAARQEGDLRAPCLYSSLPDGSWDLGLLSCVROAIADEQKV 1045

RESULT 3
US-10-032-159A-8
; Sequence 8, Application US/10032159A
; Patent No. US20020164703A1
; GENERAL INFORMATION:
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,
; TITLE OF INVENTION: ENCODING NUCLEIC ACIDS, AND METHODS OF USE
; FILE REFERENCE: P-LJ 5100
; CURRENT APPLICATION NUMBER: US/10/032,159A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,457
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-159A-8

Query Match          26.1%; Score 242.5; DB 9; Length 1247;
Best Local Similarity 31.5%; Pred. No. 1.7e-16;
Matches 57; Conservative 41; Mismatches 76; Indels 7; Gaps 4;

QY 1 RRPVLLVPRVKGKILSEKLCILQGF---FKKCLAEYLSOEYEAWSQRGDIIOEGV 55
Db 1066 RRPVLPFTPVAKTLVQRLNLSGGAMEFTICKSDIVTRDEFLRRQKTEFIYSREKNPN 1125

QY 56 GRCWVTRHAVESLMKNTALLDVQLDSVCTLHRMDIFFPIVHVSNEKMAKLLKGLQ 115
Db 1126 AFEC-IAPANIEAANKKCLLEAGTCGRDLKSNYIPVILFIRVCEKNIRFRKLLP 1184

QY 116 RLGTSEQLLEAARQEGDLRAPCLYSSLPDGSWDLGLLSCVROAIADEQKVWTE 175
Db 1185 RPET-BEFLRVCLKEKEALPCLYATVPEPDWGSVEELLRVVKDKICEBOKTIWVD 1243
```

```

QY 176 Q 176
Db 1244 E 1244

RESULT 4
US-10-029-905-12
; Sequence 12, Application US/10029905
; Patent No. US20020160438A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim Pharma KG
; TITLE OF INVENTION: Method for identifying compounds which positively
; TITLE OF INVENTION: Influence inflammatory conditions
; FILE REFERENCE: 1/1177
; CURRENT APPLICATION NUMBER: US/10/029,905
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US 60/257,854
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-029-905-12

Query Match          12.8%; Score 119.5; DB 9; Length 197;
Best Local Similarity 25.5%; Pred. No. 5.6e-05;
Matches 49; Conservative 38; Mismatches 78; Indels 27; Gaps 7;

QY 2 PRPVLLV-PRVGVKILSEKLCILQ-----GFKKCLAE-----YLSOEYEA 41
Db 4 PRPVLLSGSPGAGKSTLLKRLQLQEHSGIFGFSVSHTRNPRPGEENGKDYFVTVREVMQR 63

QY 42 WSORGDIIOEGEVSGRCWVTRHAVESLMKNTALLDVQLDSVCTLHRMDIFFPIVHVS 101
Db 64 DIAAGDFIEHAEFSGNLYCTSKVAQVQVAMNRICVLDVLDQGVNRNIKATDLRP--IYIS 121

QY 102 VNEKMAKLLKGLQRLGT-SEQLLE--AARQEGDLRAPCLYSSLPDGSWDLGLLS 158
Db 122 VQPPSLHVLEQLRQRNTETTESLVKRLAAQAADMESSKEPGLFDVWIIN--DSLDAQVA 179

QY 159 CVRQAIADQKK 170
Db 180 ELKEALSEEIKK 191

RESULT 5
US-09-919-497-98
; Sequence 98, Application US/09919497
; Patent No. US20020105662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 98
; LENGTH: 1736
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-919-497-98

Query Match          11.1%; Score 103.5; DB 10; Length 1736;
Best Local Similarity 21.2%; Pred. No. 0.044;
Matches 35; Conservative 36; Mismatches 53; Indels 41; Gaps 6;

QY 32 EYLSOE---YE-AWSQRGDIIOEGEVSGRCWVTRHAVESLMKNTALLDVQLDSVCT 87
```

Db 647 EKLAREEPIVQIAKSEPRDAGTDQRSSG-----YIRLHTIKOIIDODKHALLDVTNPNAVDR 703

Qy 88 LHRMDIFPIVHVSNKRM-----AKLKKGLORLGTSEBQLLEAARQE 131

Db 704 LNYAQWPIVVFNLNPDQKQGVKTRMRMLCPESRKSARKLYERSHKLAKNNHHLETTT--- 760

Qy 132 EGDLDRAPCLYSSLAPDQSLDGLLSCVROAIADQKKVVMTEQ 176

Db 761 -----INLNSMNDGW-----YGALKEAVQQQNQLVWVSE 790

RESULT 6

US-09-919-497-59

; Sequence 59, Application US/09919497

; Patent No. US20020106662A1

; GENERAL INFORMATION:

; APPLICANT: Mutter, George L.

; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER

; FILE REFERENCE: B0801/7225

; CURRENT APPLICATION NUMBER: US/09/919,497

; CURRENT FILING DATE: 2001-07-31

; PRIOR APPLICATION NUMBER: US 60/221,735

; PRIOR FILING DATE: 2000-07-31

; NUMBER OF SEQ ID NOS: 100

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 59

; LENGTH: 767

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-919-497-59

Query Match 9.0%; Score 84; DB 10; Length 767;

Best Local Similarity 19.2%; Pred. No. 1.5;

Matches 40; Conservative 37; Mismatches 81; Indels 50; Gaps 8;

Qy 3 RPVLVY----PRAVKILSEKLLQGFKKCL-----AEY-----LSQEEYEA 41

Db 578 RPIILGPTKDRANDDLLE---FPDKFGSCVPHTRPKREYIDGRDYHFVSSREKMEK 634

Qy 42 WSRGDIIOEGEVSGRCWVTRHAVESLMEKNTHALDVLDSVCTLHRMDIFPIVTHV- 100

Db 635 DIQAHKFEAGQYNSHLYGTSVQSVREVAEQKHCILDVSANAVRRIOAAHLHPAIFIR 694

Qy 101 -----SVNEKMA-KKLKGLORLGTSEBQLLEAARQEGLDRAPCLYSSLAPDQWS 151

Db 695 PRSLENVLEINKRITEEQARKAFDRATKLEQFTE-----CFSALVEGDSFE 741

Qy 152 DLGLLSCVROAIADQKKVVMTEQSPR 179

Db 742 E---IYHKVKRVIEDLSGPIYVWPARER 766

RESULT 7

US-09-807-721-2

; Sequence 2, Application US/09807721

; Patent No. US20020174453A1

; GENERAL INFORMATION:

; APPLICANT: AUBURN UNIVERSITY

; APPLICANT: UNIVERSITY OF CENTRAL FLORIDA

; TITLE OF INVENTION: PRODUCTION OF ANTIBODIES IN TRANSGENIC PLASTIDS

; FILE REFERENCE: 1463-PCT-US-00

; CURRENT APPLICATION NUMBER: US/09/807,721

; CURRENT FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: PCT/US01/06274

; PRIOR FILING DATE: 2001-02-28

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 669

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-807-721-2

Query Match 8.6%; Score 80; DB 9; Length 669;

Best Local Similarity 25.7%; Pred. No. 3.1;

Matches 39; Conservative 19; Mismatches 44; Indels 50; Gaps 9;

Qy 1 RRPVLLVPRAVGKILSEKL---CLLQGF--KKCLAEYLS-----QEEYEAWSQRGDII 49

Db 328 RPEVHLLPPPSEELALNELVLTCLARGFSPKDVLRWLOGSOELPREKYLTMASROEPS 387

Qy 50 Q-----EGEVSGRCWV-----TRHAVESLMEKNTHALLDV- 80

Db 388 QGTTTYAVTSILRVAAEDWKKGETFS--CMVGHEALPLAFTQKTIDRLAGKPTHINVSVV 445

Qy 81 --QLDSVCTLHRMDIF-----PIVHVSVNEKM 106

Db 446 MAEADGTC--YRMDIVMTQSPAINASPSGEKV 475

RESULT 8

US-09-895-913A-358

; Sequence 358, Application US/09895913A

; Patent No. US20020160456A1

; GENERAL INFORMATION:

; APPLICANT: Kleanthous, Harold

; APPLICANT: Al-Garawi, Amal

; APPLICANT: Miller, Charles

; APPLICANT: Tomb, Jean Francois

; APPLICANT: Oomen, Raymond P.

; TITLE OF INVENTION: Identification of Polynucleotides

; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in t

; TITLE OF INVENTION: Genome

; FILE REFERENCE: 06132/043002

; CURRENT APPLICATION NUMBER: US/09/895,913A

; CURRENT FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 08/881,227

; PRIOR FILING DATE: 1997-06-24

; NUMBER OF SEQ ID NOS: 368

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 358

; LENGTH: 867

; TYPE: PRT

; ORGANISM: Helicobacter pylori

US-09-895-913A-358

Query Match 8.4%; Score 78.5; DB 9; Length 867;

Best Local Similarity 22.0%; Pred. No. 6.3;

Matches 39; Conservative 28; Mismatches 63; Indels 47; Gaps 7;

Qy 6 LLVPRAVGKILSEKLLQGFKKCLAEYLSQEEYEAWSQRGDIIQEGEVSGRCWVTRHA 65

Db 126 LLTPEA-KKLLEAKESLKAYKDCLSQARNEE-----RRA 160

Qy 66 VESLMEKNTHALLDVQLDSVCTLHRMDIFPIVHVSVNEKMAKKLKKGLRLGTSE---- 121

Db 161 CEKLLTPEARKLLEQEVKKSIKAY-LD-----CVSRARNEKEKKECEKLLTPEARKF 211

Qy 122 --EOLLEAARQEGDLDRAPCLYSSLAPD-----GWSDDLGLLSCVROAIADQKK 170

Db 212 LAKQVNLNCKEAGNEERKACL-KNLPKDLQENILAKESLKAYKDCLSQARNEEBRR 267

RESULT 9

US-09-783-320-26

; Sequence 26, Application US/09783320

; Patent No. US20020038011A1

; GENERAL INFORMATION:

; APPLICANT: Walke, D. Wade

; APPLICANT: Hu, Yi

; APPLICANT: Nepomnichy, Boris

; APPLICANT: Turner, C. Alexander Jr

; APPLICANT: Zambrowicz, Brian

; TITLE OF INVENTION: No. US20020038011A1 Human Kinases and Polynucleotides Encodi

; FILE REFERENCE: LEX-0137-USA

; CURRENT APPLICATION NUMBER: US/09/783,320

```
; CURRENT FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: US 60/183,582
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 60/184,014
; PRIOR FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 155
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-783-320-26

Query Match
Best Local Similarity 24.5%; Score 77; DB 10; Length 155;
Matches 25; Conservative 21; Mismatches 46; Indels 10; Gaps 2;

Qy 33 YLSQEEYEAWSQGDIIQEGEVSGRCWVTRHAVESLMKNTALLDVLDSVCTLHRMD 92
Db 54 FISKHLFETDVQNNKFIEYGEYKNNYGTSDSVRSVLAKNKVCLLDVQPHTVKHLRTLE 113
Qy 93 IFPIVIVHSVNEKMAKLLKGLQRL-GTSEEQLLEAARQEEG 133
Db 114 FKPYVIFIK-----PPSIERLRRETRKNAKIISRDDQG 146

RESULT 10
US-09-783-320-28
; Sequence 28, Application US/09783320
; Patent No. US20020038011A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hu, Yi
; APPLICANT: Nepomnichy, Boris
; APPLICANT: Turner, C. Alexander Jr
; APPLICANT: Zambrowicz, Brian
; TITLE OF INVENTION: No. US20020038011A1el Human Kinases and Polynucleotides Encoding
; FILE REFERENCE: LEX-0137-USA
; CURRENT APPLICATION NUMBER: US/09/783,320
; CURRENT FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: US 60/183,582
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 60/184,014
; PRIOR FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 184
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-783-320-28

Query Match
Best Local Similarity 8.3%; Score 77; DB 10; Length 184;
Matches 25; Conservative 21; Mismatches 46; Indels 10; Gaps 2;

Qy 33 YLSQEEYEAWSQGDIIQEGEVSGRCWVTRHAVESLMKNTALLDVLDSVCTLHRMD 92
Db 83 FISKHLFETDVQNNKFIEYGEYKNNYGTSDSVRSVLAKNKVCLLDVQPHTVKHLRTLE 142
Qy 93 IFPIVIVHSVNEKMAKLLKGLQRL-GTSEEQLLEAARQEEG 133
Db 143 FKPYVIFIK-----PPSIERLRRETRKNAKIISRDDQG 175

RESULT 11
US-09-783-320-34
; Sequence 34, Application US/09783320
; Patent No. US20020038011A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hu, Yi
; APPLICANT: Nepomnichy, Boris
```

```
; APPLICANT: Turner, C. Alexander Jr
; APPLICANT: Zambrowicz, Brian
; TITLE OF INVENTION: No. US20020038011A1el Human Kinases and Polynucleotides Encodi
; FILE REFERENCE: LEX-0137-USA
; CURRENT APPLICATION NUMBER: US/09/783,320
; CURRENT FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: US 60/183,582
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 60/184,014
; PRIOR FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 195
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-783-320-34

Query Match
Best Local Similarity 8.3%; Score 77; DB 10; Length 195;
Matches 25; Conservative 21; Mismatches 46; Indels 10; Gaps 2;

Qy 33 YLSQEEYEAWSQGDIIQEGEVSGRCWVTRHAVESLMKNTALLDVLDSVCTLHRMD 92
Db 54 FISKHLFETDVQNNKFIEYGEYKNNYGTSDSVRSVLAKNKVCLLDVQPHTVKHLRTLE 113
Qy 93 IFPIVIVHSVNEKMAKLLKGLQRL-GTSEEQLLEAARQEEG 133
Db 114 FKPYVIFIK-----PPSIERLRRETRKNAKIISRDDQG 146

RESULT 12
US-09-783-320-42
; Sequence 42, Application US/09783320
; Patent No. US20020038011A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hu, Yi
; APPLICANT: Nepomnichy, Boris
; APPLICANT: Turner, C. Alexander Jr
; APPLICANT: Zambrowicz, Brian
; TITLE OF INVENTION: No. US20020038011A1el Human Kinases and Polynucleotides Encodi
; FILE REFERENCE: LEX-0137-USA
; CURRENT APPLICATION NUMBER: US/09/783,320
; CURRENT FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: US 60/183,582
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 60/184,014
; PRIOR FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 211
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-783-320-42

Query Match
Best Local Similarity 8.3%; Score 77; DB 10; Length 211;
Matches 25; Conservative 21; Mismatches 46; Indels 10; Gaps 2;

Qy 33 YLSQEEYEAWSQGDIIQEGEVSGRCWVTRHAVESLMKNTALLDVLDSVCTLHRMD 92
Db 54 FISKHLFETDVQNNKFIEYGEYKNNYGTSDSVRSVLAKNKVCLLDVQPHTVKHLRTLE 113
Qy 93 IFPIVIVHSVNEKMAKLLKGLQRL-GTSEEQLLEAARQEEG 133
Db 114 FKPYVIFIK-----PPSIERLRRETRKNAKIISRDDQG 146

RESULT 13
US-09-783-320-36
; Sequence 36, Application US/09783320
```

```

; Patent No. US20020038011A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hu, Yi
; APPLICANT: Nepomnichy, Boris
; APPLICANT: Turner, C. Alexander Jr
; APPLICANT: Zambrowicz, Brian
; TITLE OF INVENTION: No. US20020038011A1el Human Kinases and Polynucleotides Encoding
; FILE REFERENCE: LEX-0137-USA
; CURRENT APPLICATION NUMBER: US/09/783,320
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: US 60/183,582
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 60/184,014
; PRIOR FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 224
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-783-320-36

Query Match      8.3%; Score 77; DB 10; Length 224;
Best Local Similarity 24.5%; Pred. No. 1.5;
Matches 25; Conservative 21; Mismatches 46; Indels 10; Gaps 2;

QY 33 YLSQEEYEAWSORGDIIOGEYSGGRCWVTRHAVESLMKKNTHALLDVOLDVSVCTLHRMD 92
Db 83 FISKHLFETDVONNKFEIYGEYKNNYGTSDSVRSVLAKNVCLLDVQPHTVKHLRTLE 142

QY 93 IFPIVIVHSVNEKMAKLLKGLQRL-GTSEEQLLEAARQEEG 133
Db 143 FKPYVIFIK-----PPSIERLRRETRKNAKIISRDDQG 175

RESULT 14
US-09-783-320-44
; Sequence 44, Application US/09783320
; Patent No. US20020038011A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hu, Yi
; APPLICANT: Nepomnichy, Boris
; APPLICANT: Turner, C. Alexander Jr
; APPLICANT: Zambrowicz, Brian
; TITLE OF INVENTION: No. US20020038011A1el Human Kinases and Polynucleotides Encoding
; FILE REFERENCE: LEX-0137-USA
; CURRENT APPLICATION NUMBER: US/09/783,320
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: US 60/183,582
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 60/184,014
; PRIOR FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 240
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-783-320-44

Query Match      8.3%; Score 77; DB 10; Length 240;
Best Local Similarity 24.5%; Pred. No. 1.6;
Matches 25; Conservative 21; Mismatches 46; Indels 10; Gaps 2;

QY 33 YLSQEEYEAWSORGDIIOGEYSGGRCWVTRHAVESLMKKNTHALLDVOLDVSVCTLHRMD 92
Db 83 FISKHLFETDVONNKFEIYGEYKNNYGTSDSVRSVLAKNVCLLDVQPHTVKHLRTLE 142

QY 93 IFPIVIVHSVNEKMAKLLKGLQRL-GTSEEQLLEAARQEEG 133
Db 143 FKPYVIFIK-----PPSIERLRRETRKNAKIISRDDQG 175
```

```

RESULT 15
US-09-783-320-32
; Sequence 32, Application US/09783320
; Patent No. US20020038011A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hu, Yi
; APPLICANT: Nepomnichy, Boris
; APPLICANT: Turner, C. Alexander Jr
; APPLICANT: Zambrowicz, Brian
; TITLE OF INVENTION: No. US20020038011A1el Human Kinases and Polynucleotides Encodi
; FILE REFERENCE: LEX-0137-USA
; CURRENT APPLICATION NUMBER: US/09/783,320
; CURRENT FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: US 60/183,582
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 60/184,014
; PRIOR FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 296
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-783-320-32

Query Match      8.3%; Score 77; DB 10; Length 296;
Best Local Similarity 24.5%; Pred. No. 2.1;
Matches 25; Conservative 21; Mismatches 46; Indels 10; Gaps 2;

QY 33 YLSQEEYEAWSORGDIIOGEYSGGRCWVTRHAVESLMKKNTHALLDVOLDVSVCTLHRMD 92
Db 195 FISKHLFETDVONNKFEIYGEYKNNYGTSDSVRSVLAKNVCLLDVQPHTVKHLRTLE 254

QY 93 IFPIVIVHSVNEKMAKLLKGLQRL-GTSEEQLLEAARQEEG 133
Db 255 FKPYVIFIK-----PPSIERLRRETRKNAKIISRDDQG 287

Search completed: January 22, 2003, 08:52:55
Job time : 8.21774 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 13.7137 Seconds
(without alignments)
1254.807 Million cell updates/sec

Title: US-09-767-215-2_COPY_826_1004
Perfect score: 930
Sequence: 1 RPRVLLVPRVAGKILSEK.....VRQAIADQKVVWTEQSPR 179
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	130.5	14.0	1281	2 T00346	hypothetical prote
2	122.5	13.2	198	2 S39447	guanylate kinase (
3	119.5	12.8	197	2 S58864	guanylate kinase (
4	115.5	12.4	197	1 KIPG60	guanylate kinase (
5	110.5	11.9	1367	2 T13703	tama protein - fru
6	103.5	11.1	1736	2 A47747	tight junction pro
7	100.5	10.8	1745	2 A46431	tight junction-ass
8	94.5	10.2	1034	2 T22166	hypothetical prote
9	89.5	9.6	775	2 I46236	tight junction pro
10	88	9.5	1163	2 J60366	tight junction pro
11	86.5	9.3	1290	2 T00018	period protein hom
12	86	9.2	886	2 H69378	conserved hypotet
13	84.5	9.1	429	2 F82679	conserved hypotet
14	84	9.0	724	2 JH0800	postsynaptic densi
15	84	9.0	767	2 T09599	postsynaptic densi
16	83.5	9.0	646	2 T02643	hypothetical prote
17	83	8.9	911	2 I56552	synapse-associated
18	82	8.8	849	1 S44732	scaffold attachmen
19	82	8.8	1225	2 A49454	chromosome segrega
20	80.5	8.7	1047	2 G90684	ATP-dependent gsdN
21	80.5	8.7	1047	2 C85535	ATP-dependent gsdN
22	80	8.6	414	2 C69530	3-ketoacyl-CoA thi
23	80	8.6	720	2 A45436	synapse-associated
24	79.5	8.5	920	2 C96831	hypothetical prote
25	79	8.5	1744	2 F86161	F1003.10 protein -
26	78.5	8.4	379	2 T10588	UPPGlucose 4-epime
27	78.5	8.4	1920	2 A53188	pericentrin - mous
28	78.5	8.4	1927	2 G64585	csg pathogenicity
29	78	8.4	467	2 A57627	p55 erythrocyte me

30 78 8.4 814 2 D82101 oxidoreductase, ac
31 78 8.4 852 2 T10811 channel associated
32 78 8.4 1023 2 A59431 KIAA0013 protein I
33 77.5 8.3 471 2 T02749 G1/S transition co
34 77.5 8.3 1116 2 I54378 gene X104 protein
35 77 8.3 170 2 T48362 hypothetical prote
36 77 8.3 239 2 I39845 orotidine-5'-phosp
37 77 8.3 258 2 A54907 casein kinase II (c
38 77 8.3 585 2 G02165 human homolog of D
39 77 8.3 1203 2 H87687 helicase, UvrD/Rep
40 76.5 8.2 191 2 G82920 guanylate kinase U
41 76.5 8.2 375 2 T00467 probable UDPglucos
42 76.5 8.2 406 2 B35878 class I major hist
43 76.5 8.2 748 2 T02560 cellulose synthase
44 76.5 8.2 866 2 A11720 endopeptidase Clp
45 76.5 8.2 1013 2 T30818 hsdR protein - Kle

ALIGNMENTS

RESULT 1
T00346
hypothetical protein KIAA0583 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00346
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
A:Reference number: Z14086; MUID:98290545; PMID:9628581
A:Accession: T00346
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1281 <NAG>
A:Cross-references: EMBL:AB011155; NID:g3043689; PIDN:BAJ35509.1; PID:g3043690
A:Experimental source: brain; clone HJ0729
C:Genetics:
A:Note: KIAA0583
C:Superfamily: guanylate kinase homology
F:1102-1270/Domain: guanylate kinase homology <GKI>

Query Match 14.0%; Score 130.5; DB 2; Length 1281;
Best Local Similarity 26.7%; Pred. No. 0.0018;
Matches 55; Conservative 31; Mismatches 63; Indels 57; Gaps 10;

Qy 3 RPVL-----LVPRVVGK-----ILSEKLLQLOGFKKCLAEYLSQEEYAW 42
Db 1102 RPVLFLGLDVKVEMLVNEAPGKFCRCPLEYMKASQAQAIERGVRDCLF-----VDYKRR 1156
Qy 43 SORGDIQEGEVSGRCWVTRHAVESLMEKNTHALLDVQDSVCTLHRMDIFFPIVHVSU 102
Db 1157 SGHFDV-----TTVASIKETEKNRHCLLDTAHAERLHHMHIYPIVIF--I 1202
Qy 103 NEKMAKKLK-----KGLRLGTSEQLLEAARQEGDRLAPCLYSAPDGMSD 152
Db 1203 HYKSAKHIKEQRDPYLRDKVTQR--HSKEQ-FAEAQKLEQEYSR---YFTGVIOGG--A 1254
Qy 153 LDGLLSCVRQAIADQKKVWTEQSP 178
Db 1255 LSSICTQILAMVNVQKQKVLWIPACP 1280

RESULT 2
S39447
guanylate kinase (EC 2.7.4.8) - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 07-Oct-1994 #sequence_revision 05-May-1995 #text_change 19-Jan-2001
C:Accession: S39447; S39448
R:Gaidarov, I.O.; Suslov, O.N.; Abdulaev, N.G.
FEBS Lett. 335, 81-84, 1993
A:Title: Enzymes of the cyclic GMP metabolism in bovine retina. I. Cloning and expres
A:Reference number: S39448; MUID:94063086; PMID:8243671

C:Accession: F82679
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: F82679
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-429 <STM>
A:Cross-references: GB:A8003975; GB:A8003849; NID:g9106468; PIDN:AAF84262.1; GSPDB:GN001
A:Experimental source: Strain 9a5C
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1453
C:Superfamily: Haemophilus influenzae conserved hypothetical protein HT1590
Query Match 9.1%; Score 84.5; DB 2; Length 429;
Best Local Similarity 24.1%; Pred. No. 6.8;
Matches 34; Conservative 28; Mismatches 58; Indels 21; Gaps 4;
Qy 34 LSQEEYANSQGDIIQGEVSGGRCWVTHAVESLMKNTALLDVLDSVCTLHRMDI 93
Db 102 MQQDAFLPHIERGSIIFVGATT-----ENPSFELNSALLSRCRVHVL 145
Qy 94 FPIVTHSVNKKMAKKLKL--QRLGTSEQLLEAARQEGDLDRAPCLYSSSLAPDQWS 151
Db 146 VSSQDIVVALQALQDTERGCEQKIEVSEASLLLEAKAAGDVPRATLL-EIAAEALQ 204
Qy 152 DLGGLLSC--VRQATADEQKK 170
Db 205 DBDQMTADLLSQVLADRSRR 225
RESULT 14
JH0800
postsynaptic density protein PSD-95 - rat
N:Alternate names: brain specific PSD-95 protein; discs-large tumor suppressor protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C:Accession: JH0800; S26407
R:Cho, K.O.; Hunt, C.A.; Kennedy, M.B.
Neuron 9, 929-942, 1992
A:Title: The rat brain postsynaptic density fraction contains a homolog of the drosophil
A:Reference number: JH0800; MUID:93040233; PMID:1419001
A:Accession: JH0800
A:Molecule type: mRNA
A:Residues: 1-724 <CHO>
A:Cross-references: GB:M96853; NID:g206454; PIDN:AAA41971.1; PID:g206455
A:Experimental source: brain
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase hom
F:70-148/Domain: GLGF domain homology <GLG1>
F:165-243/Domain: GLGF domain homology <GLG2>
F:435-493/Domain: SH3 homology <SH3>
F:535-712/Domain: guanylate kinase homology <GKI>
Query Match 9.0%; Score 84; DB 2; Length 724;
Best Local Similarity 19.2%; Pred. No. 14;
Matches 40; Conservative 37; Mismatches 81; Indels 50; Gaps 8;

Qy 3 RPVLIV----PRAVGKILSEKLCLLQGFKKCL-----AEY-----LSQEEYEA 41
Db 535 RPIILGPTKORANDLLSE---FPDKFGSCVPHTTRPKREYEIDGRDYHFVSSREKMEK 591
Qy 42 WSQRGDIIQGEVSGGRCWVTHAVESLMKNTALLDVLDSVCTLHRMDIFPIVHV- 100
Db 592 DTQAHKFTIAGQYNHLYGTSVQSVREVAEGKHCILDVSANAVRRLOAAHLHPITAFIR 651
Qy 101 -----SVNEKMA-KKLKGLQLRGTSEQLLEAARQEGDLDRAPCLYSSSLAPDQWS 151
Db 652 PRSLENVLEINKRITEQARAFDRATKLEQFTE-----CFSIAVEGDSFE 698
Qy 152 DLGGLLSCVROAIADEQKKVWVTEQSPR 179
Db 699 E---IYHKVKRVIEDLSGPIYVWPAPER 723
RESULT 15
T09599
postsynaptic density protein 95 - human
C:Species: Homo sapiens (man)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
C:Accession: T09599
R:Stathakis, D.G.; Hoover, K.H.; You, Z.; Bryant, P.J.
submitted to the EMBL Data Library, July 1998
A:Reference number: Z16761
A:Accession: T09599
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-767 <STA>
A:Cross-references: EMBL:U83192; NID:g3318652; PID:g3318653
A:Experimental source: mammary
C:Genetics:
A:Gene: PSD95
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
F:208-286/Domain: GLGF domain homology <GLG>
F:478-536/Domain: SH3 homology <SH3>
F:578-755/Domain: guanylate kinase homology <GKI>
Query Match 9.0%; Score 84; DB 2; Length 767;
Best Local Similarity 19.2%; Pred. No. 15;
Matches 40; Conservative 37; Mismatches 81; Indels 50; Gaps 8;
Qy 3 RPVLIV----PRAVGKILSEKLCLLQGFKKCL-----AEY-----LSQEEYEA 41
Db 578 RPIILGPTKORANDLLSE---FPDKFGSCVPHTTRPKREYEIDGRDYHFVSSREKMEK 634
Qy 42 WSQRGDIIQGEVSGGRCWVTHAVESLMKNTALLDVLDSVCTLHRMDIFPIVHV- 100
Db 635 DTQAHKFTIAGQYNHLYGTSVQSVREVAEGKHCILDVSANAVRRLOAAHLHPITAFIR 694
Qy 101 -----SVNEKMA-KKLKGLQLRGTSEQLLEAARQEGDLDRAPCLYSSSLAPDQWS 151
Db 695 PRSLENVLEINKRITEQARAFDRATKLEQFTE-----CFSIAVEGDSFE 741
Qy 152 DLGGLLSCVROAIADEQKKVWVTEQSPR 179
Db 742 E---IYHKVKRVIEDLSGPIYVWPAPER 766
Search completed: January 22, 2003, 08:54:45
Job time : 16.7137 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 08:49:35 ; Search time 6.97715 Seconds
(without alignments)
1064.082 Million cell updates/sec

Title: US-09-767-215-2_COPY_826_1004

Perfect score: 930

Sequence: 1 RPRVLLVPRVKGILSEKL.....VRQAIADKQKVVWVTEQSPR 179

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	930	100.0	1004	1 CARE_HUMAN	Q9b16 homo sapien
2	738.5	79.4	999	1 CARE_MOUSE	Q99kf0 mus musculu
3	242.5	26.1	1147	1 CARB_HUMAN	Q9b17 homo sapien
4	185.5	19.9	1021	1 CARA_MOUSE	P58660 mus musculu
5	169.5	18.2	1032	1 CARA_HUMAN	Q9bwt7 homo sapien
6	124.5	13.4	197	1 KGUA_MOUSE	Q64520 mus musculu
7	122.5	13.2	197	1 KGUA_BOVIN	P46195 bos taurus
8	119.5	12.8	196	1 KGUA_HUMAN	Q16774 homo sapien
9	115.5	12.4	197	1 KGUA_PIG	P31006 sus scrofa
10	103.5	11.1	1736	1 Z01_HUMAN	Q07157 homo sapien
11	100.5	10.8	1745	1 Z01_MOUSE	P39447 mus musculu
12	93.5	10.1	1167	1 Z02_MOUSE	Q920u1 mus musculu
13	91	9.8	817	1 DLG3_HUMAN	Q92796 homo sapien
14	90	9.7	849	1 DLG3_MOUSE	P70175 mus musculu
15	90	9.7	849	1 DLG3_RAT	Q92936 rattus norv
16	89.5	9.6	1174	1 Z02_CANFA	Q95168 canis fami
17	88.5	9.5	1190	1 Z02_HUMAN	Q9udy2 homo sapien
18	88	9.5	898	1 Z03_CANFA	Q62683 canis fami
19	86.5	9.3	1290	1 PER1_HUMAN	O15534 homo sapien
20	86	9.2	886	1 RA50_ARCFU	Q92930 archaeglob
21	84	9.0	724	1 DLG4_MOUSE	Q62108 mus musculu
22	84	9.0	724	1 DLG4_RAT	P10116 rattus norv
23	84	9.0	767	1 DLG4_HUMAN	P78352 homo sapien
24	83	8.9	399	1 MTFE_HUMAN	Q99551 homo sapien
25	83	8.9	911	1 DLG1_RAT	Q62696 rattus norv
26	82	8.8	1225	1 SMCL1_YEAST	P32908 saccharomyc
27	81.5	8.8	221	1 KGUA_RAISO	Q8xxf9 ralstonia s
28	81	8.7	905	1 Z03_MOUSE	Q9qxy1 mus musculu
29	80	8.6	3210	1 CENF_HUMAN	P49454 homo sapien
30	79	8.5	933	1 Z03_HUMAN	Q95049 homo sapien
31	78.5	8.4	1920	1 PCNT_MOUSE	P48725 mus musculu
32	78	8.4	467	1 EM55_FUGRU	P49697 fugu rubrip
33	78	8.4	852	1 DLG2_RAT	Q63622 rattus norv

RESULT 1

ID	CARE_HUMAN	STANDARD:	PRT:	1004 AA.
AC	Q9BXL6; Q9BVB5;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Caspase recruitment domain protein 14 (CARD-containing MAGUK protein 2) (Carma 2).			
GN	CARD14 OR CARMA2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21192234; PubMed=11278692;			
RA	Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,			
RA	Srinivasula S.M., Merriam S., Distefano P.S., Alnemri E.S.;			
RT	"CARD14 and CARD14 are novel caspase recruitment domain			
RT	(CARD)/membrane-associated guanylate kinase (MAGUK) family members			
RT	that interact with Bcl10 and activate NF-kappaB.";			
RL	J. Biol. Chem. 276:11877-11882(2001).			
[2]				
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21255663; PubMed=11356195;			
RA	Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;			
RT	"Carmal, a CARD-containing binding partner of Bcl10, induces Bcl10			
RT	phosphorylation and NF-kappaB activation.";			
RL	FEBS Lett. 496:121-127(2001).			
[3]				
RP	ERRATUM.			
RA	Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;			
RL	FEBS Lett. 505:198-198(2001).			
[4]				
RP	SEQUENCE OF 1-740 FROM N.A.			
RC	TISSUE=Cervix, and Colon;			
RA	Strausberg R.;			
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.			
CC	!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the			
CC	phosphorylation of Bcl10.			
CC	!- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD			
CC	interaction.			
CC	!- SUBCELLULAR LOCATION: Cytoplasmic.			
CC	!- TISSUE SPECIFICITY: Expressed in placenta. Also detected in HeLa			
CC	S3 cells, but not in the other cancer cell lines tested.			
CC	!- SIMILARITY: CONTAINS 1 CARD DOMAIN.			
CC	!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.			
CC	!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.			
CC	!- CAUTION: Ppam or SMART.			
CC	!- CAUTION: Supposed to contain a SH3 domain which is not detected by			
CC	PROSITE, Pfam or SMART.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			

ALIGNMENTS

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC CC EMBL; AF322642; AAG53403.1; -.
CC CC EMBL; AY032927; AAK54453.1; -.
CC CC EMBL; BC018142; AAH18142.1; -.
CC CC EMBL; BC001326; AAH01326.1; ALT_INIT.
CC CC InterPro; IPR000619; Guanylate_kin.
CC CC InterPro; IPR001478; PDZ.
CC CC SMART; SM00072; GuK; 1.
CC CC SMART; SM00228; PDZ; 1.
CC CC PROSITE; PS02029; CARD; 1.
CC CC PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
CC CC PROSITE; PS00552; GUANYLATE_KINASE_2; 1.
CC CC PROSITE; PS0106; PDZ; 1.
CC CC Coiled coil.
CC CC DOMAIN 15 107 CARD.
CC CC FT DOMAIN 128 409 COILED COIL (POTENTIAL).
CC CC FT DOMAIN 568 658 PDZ.
CC CC FT DOMAIN 858 990 GUANYLATE_KINASE.
CC CC FT CONFLICT 619 671 DYASEPLFKAVLEDTTLEEAAGLLRRVDGFCCLSVKNTD
CC CC FT GYKLLQLEAK -> SRARELLSPGLMGTVAAAGVTQAD
CC CC FT FTSPRCRSTLGNWASLSWADVKKSAHL (IN REF. 2;
CC CC AAH01326).
CC CC SEQUENCE 1004 AA; 113299 MW; 7469B8B5B8E06073 CRC64;

Query Match 100.08; Score 930; DB 1; Length 1004;
Best Local Similarity 100.08; Pred. No. 2.9e-73;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRPVLLVPRAVGKILSEKLCILQGFKKLAELYLSQEEYEAWNSQSGDIITQEGVSGRCW 60
DB 826 RRPVLLVPRAVGKILSEKLCILQGFKKLAELYLSQEEYEAWNSQSGDIITQEGVSGRCW 885
QY 61 VTRHAVESLMENKTHALLDVQDSVCTLRHMDIFPIVIVHSVNEKMAKLLKGLQRLGTS 120
DB 886 VTRHAVESLMENKTHALLDVQDSVCTLRHMDIFPIVIVHSVNEKMAKLLKGLQRLGTS 945
QY 121 EQOLLEAARQEEGDLDRAPCLYSSLPADGWSLDGLLSCVROAIADEQKVVVWTEQSPR 179
DB 945 EQOLLEAARQEEGDLDRAPCLYSSLPADGWSLDGLLSCVROAIADEQKVVVWTEQSPR 1004

RESULT 2
CARE_MOUSE
ID CARE_MOUSE STANDARD; PRT; 999 AA.
AC Q99KE0.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 14 (Bcl10-interacting MAGUK protein 2) (Bimp2).
GN CARD14 OR BIMP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21391892; PubMed=11387339;
RA McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A.,
RA Li Q., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W.,
RA Nunez G.;
RT "Bimp1, a MAGUK family member linking protein kinase C activation to
RT Bcl10-mediated NF-kappa B induction."
RL J. Biol. Chem. 276:30589-30597(2001).
RN [2]
RP SEQUENCE OF 82-743 FROM N.A.
RC TISSUE=Breast;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
```

```
CC CC -!- SUBUNIT: CARD14 and Bcl10 bind to each other by CARD-CARD
CC CC interaction (By similarity).
CC CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC CC -!- SIMILARITY: CONTAINS 1 GUANYLATE_KINASE-LIKE DOMAIN.
CC CC -!- CAUTION: Supposed to contain a SH3 domain which is not detected by
CC CC PROSITE, Pfam or SMART.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
```

```
CC CC EMBL; AF363457; AAK60137.1; -.
CC CC EMBL; BC004692; AAH04692.1; -.
CC CC InterPro; IPR000619; Guanylate_kin.
CC CC InterPro; IPR001478; PDZ.
CC CC Pfam; PF00595; PDZ; 1.
CC CC PROSITE; PS02029; CARD; 1.
CC CC PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
CC CC PROSITE; PS00552; GUANYLATE_KINASE_2; 1.
CC CC PROSITE; PS0106; PDZ; 1.
CC CC Coiled coil.
CC CC DOMAIN 15 107 CARD.
CC CC FT DOMAIN 125 411 COILED COIL (POTENTIAL).
CC CC FT DOMAIN 572 655 PDZ.
CC CC FT DOMAIN 854 986 GUANYLATE_KINASE.
CC CC FT CONFLICT 736 743 QAQOOLLA -> HLEDPHRS (IN REF. 2).
CC CC SEQUENCE 999 AA; 113496 MW; D18350DA12430255 CRC64;

Query Match 79.48; Score 738.5; DB 1; Length 999;
Best Local Similarity 80.38; Pred. No. 1.3e-56;
Matches 143; Conservative 11; Mismatches 23; Indels 1; Gaps 1;

QY 1 RRPVLLVPRAVGKILSEKLCILQGFKKLAELYLSQEEYEAWNSQSGDIITQEGVSGRCW 60
DB 822 RRPVLLVPRAVGKILSEKLCILQGFKKLAELYLSQEEYEAWNSQSGDIITQEGVSGRCW 881
QY 61 VTRHAVESLMENKTHALLDVQDSVCTLRHMDIFPIVIVHSVNEKMAKLLKGLQRLGTS 120
DB 882 VTRHAVESLMENKTHALLDVQDSVCTLRHMDIFPIVIVHSVNEKMAKLLKGLQRLGTS 941
QY 121 EQOLLEAARQEEGDLDRAPCLYSSLPADGWSLDGLLSCVROAIADEQKVVVWTEQSP 178
DB 942 EQOLLEAARQEEGDLDRAPCLYSSLPADGWSLDGLLSCVROAIADEQKVVVWTE-SP 998

RESULT 3
CARE_HUMAN
ID CARE_HUMAN STANDARD; PRT; 1147 AA.
AC Q9EXL7.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 11 (CARD-containing MAGUK protein 3) (Carma 1).
GN CARD11 OR CARVAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21192234; PubMed=11278692;
RA Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA Srinivasula S.M., Merriam S., DiStefano P.S., Alnemri E.S.;
RT "CARD11 and CARD14 are novel caspase recruitment domain
RT (CARD)/membrane-associated guanylate kinase (MAGUK) family members
RT that interact with Bcl10 and activate NF-kappaB."
RN [2]
```

J. Biol. Chem. 276:11877-11882(2001).

RESULT 4

CARD_MOUSE

AC P38660; 2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Caspase recruitment domain protein 10 (Bcl10-interacting MAGUK protein 1) (Bim1).

GN CARD10 OR BIM1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21391892; PubMed=11387339;

RA McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A., Li Q., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W., Nunez G.;

RT "Bim1, a MAGUK family member linking protein kinase c activation to Bcl10-mediated NF-kappa B induction.";

RL J. Biol. Chem. 276:30589-30597(2001).

CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.

CC -!- SUBUNIT: CARD10 and Bcl10 bind to each other by CARD-CARD interaction. They both participate in a complex with MALT1, where MALT1 binds to Bcl10.

CC -!- TISSUE SPECIFICITY: Highly expressed in kidney, heart followed by brain, lung, liver, skeletal muscle and testis.

CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.

CC -!- CAUTION: Supposed to contain a SH3, a PDZ and a quanylate kinase-like domain. But none of these 3 domains are detected by PROSITE, Pfam or SMART.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AF363456; AAK60136.1; -.

DR PROSITE; PS50209; CARD; 1.

KW Coiled coil.

FT DOMAIN 23 115 CARD.

FT DOMAIN 138 450 COILED COIL (POTENTIAL).

FT DOMAIN 558 565 POLY-SER.

SQ SEQUENCE 1021 AA; 114413 MW; 4811A09BDB8F792C CRC64;

Query Match 19.98; Score 185.5; DB 1; Length 1021;

Best Local Similarity 31.08; Pred. No. 1.6e-08;

Matches 58; Conservative 25; Mismatches 79; Indels 25; Gaps 5;

Qy 3 RPYLLVPRVAVGKILSEKLCILQGG----FKKCLAEYLSQEEY-----EAWSQRGDII 49

Db 834 RPYLLVPRVAVGKILSEKLCILQGG----FKKCLAEYLSQEEY-----EAWSQRGDII 49

Qy 50 QGEVSGGRCWVTHAVESLMKNTKTHALLDVLQDSVCTLHRMDIFFPIVTHVSVNEKMAK 109

Db 893 -----GSR---IRAIQESVGGK--HCLLELCARGVRELHVHSEVPIVTHVEVTEKNVRE 941

Qy 110 LKKGLORLGTSEEQLLEAARQEGDLDRAPCLYSLAPDGSWDLGLSCVRAQADQK 169

Db 942 IRGLGRPGWDSSELLRCRGSEOWLWGLPCSWQVPAHAWGHABELAKVVRGRLQSOA 1001

Qy 170 KVVWTEQ 176

Db 1002 RLWVVER 1008

RESULT 5

J. Biol. Chem. 276:11877-11882(2001).

RESULT 4

CARD_MOUSE

AC P38660; 2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Caspase recruitment domain protein 10 (Bcl10-interacting MAGUK protein 1) (Bim1).

GN CARD10 OR BIM1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21391892; PubMed=11387339;

RA McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A., Li Q., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W., Nunez G.;

RT "Bim1, a MAGUK family member linking protein kinase c activation to Bcl10-mediated NF-kappa B induction.";

RL J. Biol. Chem. 276:30589-30597(2001).

CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.

CC -!- SUBUNIT: CARD10 and Bcl10 bind to each other by CARD-CARD interaction. They both participate in a complex with MALT1, where MALT1 binds to Bcl10.

CC -!- TISSUE SPECIFICITY: Highly expressed in kidney, heart followed by brain, lung, liver, skeletal muscle and testis.

CC -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.

CC -!- CAUTION: Supposed to contain a SH3, a PDZ and a quanylate kinase-like domain. But none of these 3 domains are detected by PROSITE, Pfam or SMART.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AF363456; AAK60136.1; -.

DR PROSITE; PS50209; CARD; 1.

KW Coiled coil.

FT DOMAIN 23 115 CARD.

FT DOMAIN 138 450 COILED COIL (POTENTIAL).

FT DOMAIN 558 565 POLY-SER.

SQ SEQUENCE 1021 AA; 114413 MW; 4811A09BDB8F792C CRC64;

Query Match 19.98; Score 185.5; DB 1; Length 1021;

Best Local Similarity 31.08; Pred. No. 1.6e-08;

Matches 58; Conservative 25; Mismatches 79; Indels 25; Gaps 5;

Qy 3 RPYLLVPRVAVGKILSEKLCILQGG----FKKCLAEYLSQEEY-----EAWSQRGDII 49

Db 834 RPYLLVPRVAVGKILSEKLCILQGG----FKKCLAEYLSQEEY-----EAWSQRGDII 49

Qy 50 QGEVSGGRCWVTHAVESLMKNTKTHALLDVLQDSVCTLHRMDIFFPIVTHVSVNEKMAK 109

Db 893 -----GSR---IRAIQESVGGK--HCLLELCARGVRELHVHSEVPIVTHVEVTEKNVRE 941

Qy 110 LKKGLORLGTSEEQLLEAARQEGDLDRAPCLYSLAPDGSWDLGLSCVRAQADQK 169

Db 942 IRGLGRPGWDSSELLRCRGSEOWLWGLPCSWQVPAHAWGHABELAKVVRGRLQSOA 1001

Qy 170 KVVWTEQ 176

Db 1002 RLWVVER 1008

RESULT 5


```

RX MEDLINE=94271265; PubMed=7911663;
DR Gaidarov I.O., Suslov O.N., Ovchinnikova T.V., Abdulaev N.G.;
RT "Guanylate kinase from bovine retina: isolation, primary structure,
RL and expression in E. coli.";
RL Bioorg. Khim. 20:367-381(1994).
CC -I- FUNCTION: ESSENTIAL FOR RECYCLING GMP AND INDIRECTLY, CGMP.
CC -I- CATALYTIC ACTIVITY: ATP + GMP -- ADP + GDP.
CC -I- SUBUNIT: MONOMER (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE GUANYLATE KINASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X67029; CAAT7423.1; -.
DR HSSP; P15454; IGKY.
DR InterPro; IPR000619; Guanylate_kin.
DR Pfam; PF00625; Guanylate_kin; 1.
DR SMART; SM00072; GuKc; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS00932; GUANYLATE_KINASE_2; 1.
KW Transferase; Kinase; ATP-binding; Acetylation.
FT INIT_MET 0 BY SIMILARITY.
FT NP_BIND 10 17 ACETYLATION (BY SIMILARITY).
FT MOD_RES 1 ATYPICALITY.
SQ SEQUENCE 197 AA; 21778 MW; 5CB5DD007BC15C62 CRC64;
Query Match 13.2%; Score 122.5; DB 1; Length 197;
Best Local Similarity 24.5%; Pred. No. 0.00076;
Matches 51; Conservative 35; Mismatches 77; Indels 45; Gaps 7;
QY 2 PRPVLLV-PRANGKILSEKLCLLOGFFKCAAE-----Y 33
Db ||||| | | | | | | | | | | | | | | | | :
3 PRPVLSGPCAGAKSTL-----LKKLLOEHGSIFGFSVSHTRDPPRGEENGKDYYF 54
QY 34 LSQEYEAWSGQRDIIOEGEVSGRCWVTTHAVESLMKEKNTHALLDVLDSVCCTLHRMDI 93
Db ::::: | | | | | | | | | | | | | | | | :
55 VTRVMORDIANGDFIEHAFESGNLYGTSAARAVALQAMNRFICVLVDLQGVNRNIKTDL 114
QY 94 FPIVIHVSVN--EKMAKKLKKGGLRGLTSEQLLE--AAROEGLDRAPCLYSLSAPDG 149
Db || | | : : : : | | | | | | | | | | | :
115 RPIVFQPPSLDVLEQRLR--QRNTETEESLAKRLAAARADSWESSKEPGLFDLIIVN- 170
QY 150 WSDLDGLSCVRQAIAEDQKVVWTEQS 177
Db || : : : : | | | | | | | | | | | |
171 -DSLDKAYWALKALSEIKKAQGTGOS 197
RESULT 8
KGUA_HUMAN
ID KGUA_HUMAN STANDARD; PRT; 196 AA.
AC Q16774;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Guanylate kinase (EC 2.7.4.8) (GMP kinase).
DN GUK1 OR GMK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=96213684; PubMed=8647247;
RA Fitzgibbon J., Katsanis N., Wells D., Delhanty J., Vallins W.,
RA Hunt D.M.;
RT "Human guanylate kinase (GUK1): cDNA sequence, expression and
RT chromosomal localisation."

```

```

RL FEBS Lett. 385:185-188(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96279248; PubMed=8663313;
RT Brady W.A., Kokoris M.S., Fitzgibbon M., Black M.E.;
RT "Cloning, characterization, and modeling of mouse and human guanylate
RT kinases.";
RL J. Biol. Chem. 271:16734-16740(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ESSENTIAL FOR RECYCLING GMP AND INDIRECTLY, CGMP.
CC -!- CATALYTIC ACTIVITY: ATP + GMP -> ADP + GDP.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GUANYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L76200; AAC37598.1; -.
DR EMBL: U66895; AAC50859.1; -.
DR EMBL: BC008249; AAH06249.1; -.
DR EMBL: BC009914; AAH09914.1; -.
DR HSSP: P15454; IGKY.
DR Genew; HGNC:4693; GUK1.
DR MIM: 139270; -.
DR InterPro: IPR000619; Guanylate_kin.
DR Pfam: PF00625; Guanylate_kin; 1.
DR SMART: SM00072; GUKC; 1.
DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR Transferase; Kinase; ATP-binding; Acetylation.
DR INIT_MET 0 0 BY SIMILARITY.
DR NP_BIND 10 17 ATP (BY SIMILARITY).
DR MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
DR SEQUENCE 196 AA; 21594 MW; C4727A/E2AA261B3 CRC64;

Query Match 12.8%; Score 119.5; DB 1; Length 196;
Best Local Similarity 25.5%; Pred. No. 0.0014;
Matches 49; Conservative 38; Mismatches 78; Indels 27; Gaps 7;

QY 2 PRPVLLV-PRAVGKILSEKLCCLLQ-----GFKKCLAE-----YLSQEEYEA 41
DB 3 PRPVLSGSPGAGKSTLLKRLQLQEHSGIFGFSVSHTRNPRGGEENGKDYIFVTREVMQR 62
QY 42 WSORGDIQEGVSGRCWTRHVESLMKNTHALLDVDSVCTLHRMDIFPIVHVS 101
DB 63 DIAAGDFIEHAFFSGNLYGTSKVAQVAVQAMNRCICVLDVLDQGVNRKATDLRP--IYIS 120
QY 102 VNEKMAKKLKGQLRGCT--SEEOLLE--AARQEGDLDRAPCLYSSLPDGSWDLGLLS 158
DB 121 VQPSLHVLEQLRQRNTETESLVRKLAQAADMESKEPGLFVVIIIN--DSLDOAYA 178
QY 159 CVRQATAEQKK 170
DB 179 ELKEALSEEIKK 190

RESULT 9
KGUA_PIG
ID KGUA_PIG STANDARD; PRT; 197 AA.
AC P31006;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Guanylate kinase (EC 2.7.4.8) (GMP kinase).

```

```

GN GUK1 OR GUK.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
ON NCBI_TaxID=9823;
RX [1]
RN SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=93238695; PubMed=8097461;
RA Zschocke P.D., Schiltz E., Schulz G.E.;
RT "Purification and sequence determination of guanylate kinase from pig
RT brain.";
RL Eur. J. Biochem. 213:263-269(1993).
CC -!- FUNCTION: ESSENTIAL FOR RECYCLING GMP AND INDIRECTLY, CGMP.
CC -!- CATALYTIC ACTIVITY: ATP + GMP = ADP + GDP.
CC -!- SUBUNIT: MONOMER.
CC -!- SIMILARITY: BELONGS TO THE GUANYLATE KINASE FAMILY.
CC -----
DR PIR: S23776; KIFGGU.
DR HSSP: P15454; IGKY.
DR InterPro: IPR000619; Guanylate_kin.
DR Pfam: PF00625; Guanylate_kin; 1.
DR SMART: SM00072; GUKC; 1.
DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR Transferase; Kinase; ATP-binding; Acetylation.
DR INIT_MET 0 0
DR NP_BIND 10 17 ATP (BY SIMILARITY).
DR MOD_RES 1 1 ACETYLATION.
DR SEQUENCE 197 AA; 21789 MW; CF492B786FCC4E6E CRC64;

Query Match 12.4%; Score 115.5; DB 1; Length 197;
Best Local Similarity 24.0%; Pred. No. 0.0031;
Matches 50; Conservative 35; Mismatches 78; Indels 45; Gaps 7;

QY 2 PRPVLLV-PRAVGKILSEKLCCLLQGGFKKCLAE-----Y 33
DB 3 PRPVLSGSPGAGKSTL-----LKKLQEHSGIFGFSVSHTRDRPPGGEENGKDYIF 54
QY 34 LSQEEYEAWSQEGDIQEGVSGRCWTRHVESLMKNTHALLDVDSVCTLHRMDI 93
DB 55 VTREVMQRDIAAGDFIEHAFFSGNLYGTSKAAVRAVQAMNRCICVLDVLDQGVNRKKTDL 114
QY 94 FPIVHVSVN--EKMAKKLKGQLRGCTSEEOLLE--AARQEGDLDRAPCLYSSLPDGP 149
DB 115 QPIYIFVQPPSLDVLQRLR--QRNTETESLAKRLAAAKADMESKEPGLFLLIIN- 170
QY 150 WSDLGLLSCVPRQATAEQKKVVTWEQS 177
DB 171 -DSLDRAYWALKEALSEEIKKAQATGHS 197

RESULT 10
ZOL_HUMAN
ID ZOL_HUMAN STANDARD; PRT; 1736 AA.
AC Q07157;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-1 (Zonula occludens 1 protein) (Zona
DE occludens 1 protein) (tight junction protein 1).
GN TJP1 OR ZOL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93361541; PubMed=8395056;
RA Willott E., Balda M.S., Fanning A.S., Jameson B., van Itallie C.,
RA Anderson J.M.;
RT "The tight junction protein ZO-1 is homologous to the Drosophila
RT discs-large tumor suppressor protein of septate junctions.";

```

```
RL Proc. Natl. Acad. Sci. U.S.A. 90:7834-7838(1993).
CC -!- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL
CC REQUIRED FOR TIGHT JUNCTION ASSEMBLY, WHILE THE C-TERMINAL MAY
CC HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN
CC MIGHT BE INVOLVED IN STABILIZING JUNCTIONS.
CC -!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-3.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE, CYTOPLASMIC SIDE.
CC MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT
CC OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: THE ALPHA-CONTAINING ISOFORM IS FOUND IN MOST
CC EPITHELIAL CELL JUNCTIONS. THE SHORT ISOFORM IS FOUND BOTH IN
CC ENDOTHELIAL CELLS AND THE HIGHLY SPECIALIZED EPITHELIAL JUNCTIONS
CC OF RENAL GLOMERULI AND SERTOLI CELLS OF THE SEMINIFEROUS TUBULES.
CC -!- PTM: PHOSPHORYLATED.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L14837; AAA02891.1; -.
CC DR A7747; A47747.
CC DR HSP; P31016; IBE9.
CC DR Genew; HGNC:11827; TJP1.
CC DR MIM; 601009; -.
CC DR InterPro; IPR000619; Guanylate_kin.
CC DR InterPro; IPR001478; PDZ.
CC DR InterPro; IPR001452; SH3.
CC DR InterPro; IPR000906; ZU5.
CC DR Pfam; PF00595; PDZ; 3.
CC DR Pfam; PF00625; Guanylate_kin; 1.
CC DR Pfam; PF00791; ZU5; 1.
CC DR SMART; SM00072; GUKC; 1.
CC DR SMART; SM00228; PDZ; 3.
CC DR SMART; SM00326; SH3; 1.
CC DR SMART; SM00218; ZU5; 1.
CC DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
CC DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
CC DR PROSITE; PS50106; PDZ; 3.
CC DR PROSITE; PS50002; SH3; 1.
CC DR PROSITE; PS50002; SH3; 1.
CC KW Tight junction; SH3 domain; Alternative splicing; Repeat; Membrane;
CC Phosphorylation.
CC FT DOMAIN 11 98 PDZ 1.
CC FT DOMAIN 174 252 PDZ 2.
CC FT DOMAIN 409 490 PDZ 3.
CC FT DOMAIN 504 572 SH3.
CC FT DOMAIN 632 782 GUANYLATE KINASE.
CC FT DOMAIN 1231 1236 POLY-PRO.
CC FT DOMAIN 1414 1420 POLY-PRO.
CC FT VARSPIC 910 989 MISSING (IN SHORT ISOFORM).
CC SEQUENCE 1736 AA; 194721 MH; 508D01B7A0814FFE CRC64;

Query Match 11.1%; Score 103.5; DB 1; Length 1736;
Best Local Similarity 21.2%; Pred. NO. 0.41;
Matches 35; Conservative 36; Mismatches 53; Indels 41; Gaps 6;

QY 32 EYLSQEE---YE-ANSQRCDIIQEGVSGRCWTRHAVESLMKNTALLDVQLDSVCT 87
DB 647 EKLAREEPIYQIAKSEPRDAGTDQRSSG---YIRLTKIQIIDQKHALLDVTPNADR 703
QY 88 LHRMDFPIVIVHSVNEKM-----AKLLKGLQRLGTSEEOLLEAARQE 131
DB 704 LNYAQWPIVWFENFDSKOGVYTMRLCPESRKARKLYERSHKLAKNNHLLFTTT--- 760
DB 704 LNYAQWPIVWFENFDSKOGVYTMRLCPESRKARKLYERSHKLAKNNHLLFTTT--- 760
```

```
QY 132 EGDLDRAFCFLYSSLPADGWSDDLGLLSCVQRQAIADQKKVYVWTEQ 176
DB 761 -----INLNSMNDGW-----YGALKEAVQQQNLQVWVSE 790

RESULT 11
ZOL_MOUSE
ID ZOL_MOUSE STANDARD; PRT; 1745 AA.
AC P39447;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-1 (Zonula occludens 1 protein) (Zona
DE occludens 1 protein) (tight junction protein 1).
GN TJP1 OR ZOI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=93252986; PubMed=8486731;
RA Itoh M., Nagafuchi A., Yonemura S., Yasuda-Kitani T., Tsukita S.,
RA Tsukita S.;
RT "The 220-kD protein colocalizing with cadherins in non-epithelial
RT cells is identical to ZO-1, a tight junction-associated protein in
RT epithelial cells: cDNA cloning and immunoelectron microscopy.";
RL J. Cell Biol. 121:491-502(1993).
CC -!- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL
CC REQUIRED FOR TIGHT JUNCTION ASSEMBLY, WHILE THE C-TERMINAL MAY
CC HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN
CC MIGHT BE INVOLVED IN STABILIZING JUNCTIONS.
CC -!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-3.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE, CYTOPLASMIC SIDE.
CC MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT
CC OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D14340; BAA03274.1; -.
CC DR PIR; A46431; A46431.
CC DR HSP; P31016; IBE9.
CC DR MGD; MGI:98759; TJP1.
CC DR InterPro; IPR000619; Guanylate_kin.
CC DR InterPro; IPR001478; PDZ.
CC DR InterPro; IPR001452; SH3.
CC DR InterPro; IPR000906; ZU5.
CC DR Pfam; PF00018; SH3; 1.
CC DR Pfam; PF00595; PDZ; 3.
CC DR Pfam; PF00625; Guanylate_kin; 1.
CC DR Pfam; PF00791; ZU5; 1.
CC DR SMART; SM00072; GUKC; 1.
CC DR SMART; SM00228; PDZ; 3.
CC DR SMART; SM00326; SH3; 1.
CC DR SMART; SM00218; ZU5; 1.
CC DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
CC DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
CC DR PROSITE; PS50106; PDZ; 3.
CC DR PROSITE; PS50002; SH3; 1.
CC KW Tight junction; SH3 domain; Repeat; Membrane.
CC FT DOMAIN 23 110 PDZ 1.
CC FT DOMAIN 186 264 PDZ 2.
```


[illegible]


```
Qy 176 Q 176
Db 1168 E 1168

RESULT 2
Q9Y4E3 PRELIMINARY; PRT; 1281 AA.
AC Q9V4E3
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE KIAA0583 protein (fragment).
GN KIAA0583
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shah G., Brugada R., Roberts R.;
RT "Complete cloning and genetic organization of KIAA0583.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF352033; AAL83937.1; -.
SQ SEQUENCE 1809 AA; 202039 MW; 4794A6C8561CD905 CRC64;

Query Match 14.0%; Score 130.5; DB 4; Length 1809;
Best Local Similarity 26.7%; Pred. No. 0.0032;
Matches 55; Conservative 31; Mismatches 63; Indels 57; Gaps 10;

Qy 3 RPVL-----LVPRVVGK-----ILSEKLCLOGFKKCLAEVLSQEEYEA 42
Db 1630 RPVLILGPLLDVVVKEMLVNEAPGKFCRCPLVNMKASQAIEGVKDCLF-----VDYKRR 1684
Qy 43 SQRGDIIEGEVSGRCWCWTRHAVESLMKENTHALLDVQLDSVCTLHRMDIFPIVHVS 102
Db 1685 SGHFDV-----TTVASIKETKRNHCLLDIAPHAIERLHHMHIYPVIF--I 1730
Qy 103 NEKMAKKL-----KGLRLGTSEEQLLEAARQEGDLDRAPCLYSSLPADGWS 152
Db 1731 HYSAKHKEQRPDIYLRDKVTQR--HSKEQ--FEAAQKLEQYSR---YFTGVIQGG--A 1782
Qy 153 LDGLLSCVRQAIADDEQKVVWTEQSP 178
Db 1783 LSSICTQILAMVNEQNKVLWIPACP 1808

RESULT 4
Q8TDM6 PRELIMINARY; PRT; 1809 AA.
AC Q8TDM6
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE DLG5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shah G., Brugada R., Roberts R.;
RT "Complete CDNA of KIAA0583/DLG5.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF352034; AAL83938.1; -.
SQ SEQUENCE 1809 AA; 202066 MW; 3B77B3DCAD6FD6E9 CRC64;

Query Match 14.0%; Score 130.5; DB 4; Length 1809;
Best Local Similarity 26.7%; Pred. No. 0.0032;
Matches 55; Conservative 31; Mismatches 63; Indels 57; Gaps 10;

Qy 3 RPVL-----LVPRVVGK-----ILSEKLCLOGFKKCLAEVLSQEEYEA 42
Db 1630 RPVLILGPLLDVVVKEMLVNEAPGKFCRCPLVNMKASQAIEGVKDCLF-----VDYKRR 1684
Qy 43 SQRGDIIEGEVSGRCWCWTRHAVESLMKENTHALLDVQLDSVCTLHRMDIFPIVHVS 102
Db 1685 SGHFDV-----TTVASIKETKRNHCLLDIAPHAIERLHHMHIYPVIF--I 1730
Qy 103 NEKMAKKL-----KGLRLGTSEEQLLEAARQEGDLDRAPCLYSSLPADGWS 152
```



```

RESULT 10
Q9VHK4 ID Q9VHK4 PRELIMINARY; PRT; 1445 AA.
AC DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE PYD protein.
GN PYD OR CG9763.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003680; AAF54300.1; -.
DR HSSP; Q12923; 3PDZ.
DR FlyBase; FBgn0003177; pyd.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR000906; ZU5.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00791; ZU5; 1.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00218; ZU5; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 6.
SQ SEQUENCE 1445 AA; 156723 MW; A5675FB66676A40E CRC64;
Query Match 11.9%; Score 110.5; DB 5; Length 1445;
Best Local Similarity 23.4%; Pred No. 0.18; Mismatches 64; Indels 43; Gaps 9;
Matches 44; Conservative 37;
Qy 3 RPVLVPRVAGKILSEKILQGFKKLAELYSQBEYPAWSQSGDIIOEGEYSGRCWYT 62
Db 632 RPVLVF-GPVSDLARERLA--KDFPKFTSTPLQDDKSA-----ATSGKCRIV 676
Qy 63 RHA-VESLMKXNTHALLDVOLDVDSVCTLHRMDIFPIVIVSVNEK-MAKKLKKGLORLG-T 119
Db 677 RLNSNRDVMGRGKHALLDITPNAVRLNYAOPYPVVIKFTDSKRVIKQLRGLPKAAHK 736
Qy 120 SEQQLLEAARQEGDLDRAPCLYSLSLAPDGW-----SDLDGLLSCVROAIADEQK 169
Db 737 SSKLLEOCOK---LERV-----WSHIFSTQIALSDSESWYRKLRDSDLDQOS 781
Qy 170 KVVWTEQS 177
Db 782 GAVWMSSES 789
RESULT 11
Q9VKG8 ID Q9VKG8 PRELIMINARY; PRT; 1916 AA.
AC DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CG6509 protein (LD32687p).
GN CG6509.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003680; AAF54300.1; -.
DR HSSP; Q12923; 3PDZ.
DR FlyBase; FBgn0003177; pyd.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR000906; ZU5.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00791; ZU5; 1.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00218; ZU5; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.

```


DR	Pfam; PF00625; Guanylate_kin; 1.
DR	Pfam; PF00595; PDZ; 1.
DR	Pfam; PF00018; SH3; 1.
DR	PRINTS; PR01597; ZONOCCLUDNS.
DR	PRINTS; PR01599; ZONOCCLUDNS2.
DR	SMART; SM00072; GukC; 1.
DR	SMART; SM00228; PDZ; 1.
DR	SMART; SM00326; SH3; 1.
DR	PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR	PROSITE; PS50106; PDZ; 1.
DR	PROSITE; PS50002; SH3; 1.
KW	SH3 domain.
FT	NON_TER .
SQ	SEQUENCE 813 AA; 90776 MW; 30EA78F19DE9C852 CRC64;
Query Match 10.7%; Score 99.5; DB 11; Length 813;	
Best Local Similarity 23.5%; Pred. No. 0.94;	
Matches 43; Conservative 38; Mismatches 63; Indels 39; Gaps 10;	
QY	18 EKLCLLQ-GFKKCLAEY-----LSQEYEAWSQRGDIQE-----GEVSGGRCWVT 62 : : : : : : : : : : : : : : :
Db	347 EKVLLREAGFKPVVLEGPADIAMERLT--TELPDLFTAQTAKPEKDAGESEKSSG--VVR 402
QY	63 RHAVESLMETKTHALLDVQLDSVCTLHRMDDIPPIVIHVSNEKVAKKLKGLRLGTSEE 122 : : : :::: : : : : : : : : : : : : : : : : : : :
Db	403 LNTVKYIIEQDKHALLDVTPKAVDLLHYTFPPIVIFFTPDSRQG--IKTIROKLNPSTN 460
QY	123 QLLEAARQ--EEGDLDRAPC-----LYSSLAPDGNISDLGLSLSCVROATADEQKKVVW 173 : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	461 KI---SRKLFDPQANKSKTCSHLTLATINVNSANDGW-----FGSLKDSIQQQQHSAVW 511
QY	174 TEQ 176
Db	512 VSE 514
RESULT 14	
Q20398	PRELIMINARY; PRT; 1034 AA.
ID	Q20398 AC Q20398;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	F44D12.1 protein.
GS	F44D12.1.
OS	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
QC	Rhabditiidae; Peloderinae; Caenorhabditis.
OX	NCHI_Taxid=6239;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Coles L.;
RL	Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=99069613; PubMed=9851916;
RA	none;
RT	"Genome sequence of the nematode C.elegans: A platform for investigating biology";
RL	Science 282:2012-2018(1998).
DR	EMBL; Z68298; CAAG2607.1; -.
DR	InterPro; IPR000619; Guanylate_kin.
DR	InterPro; IPR001478; PDZ.
DR	Pfam; PF00595; PDZ; 1.
DR	SMART; SM00072; GukC; 1.
DR	SMART; SM00228; PDZ; 3.
DR	PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR	PROSITE; PS50106; PDZ; 2.
SQ	SEQUENCE 1034 AA; 116053 MW; 7E29646264FCDB5B CRC64;
Query Match 10.2%; Score 94.5; DB 5; Length 1034;	
Best Local Similarity 25.6%; Pred. No. 3.6;	
Matches 32; Conservative 29; Mismatches 47; Indels 17; Gaps 17;	

